Astronomy
&
Astrophysics

KiDS-1000: Cosmology with improved cosmic shear measurements

Shun-Sheng Li[1], Henk Hoekstra[1], Konrad Kuijken[1], Marika Asgari[2,3], Maciej Bilicki[4], Benjamin Giblin[5,6], Catherine Heymans[6,7], Hendrik Hildebrandt[7], Benjamin Joachimi[8], Lance Miller[9], Jan Luca van den Busch[7], Angus H. Wright[7], Arun Kannawadi[10], Robert Reischke[7], and HuanYuan Shan[11,12,13]

[1] Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands
e-mail: ssli@strw.leidenuniv.nl
[2] E. A. Milne Centre, University of Hull, Cottingham Road, Hull HU6 7RX, UK
[3] Centre of Excellence for Data Science, AI, and Modelling (DAIM), University of Hull, Cottingham Road, Hull HU6 7RX, UK
[4] Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland
[5] Instituto de Ciencias del Cosmos (ICC), Universidad de Barcelona, Martí i Franquès, 1, 08028 Barcelona, Spain
[6] Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
[7] Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing, 44780 Bochum, Germany
[8] Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
[9] Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK
[10] Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544, USA
[11] Shanghai Astronomical Observatory (SHAO), Nandan Road 80, Shanghai 200030, PR China
[12] Key Laboratory of Radio Astronomy and Technology, Chinese Academy of Sciences, A20 Datun Road, Chaoyang District, Beijing 100101, PR China
[13] University of Chinese Academy of Sciences, Beijing 100049, PR China

ABSTRACT

We present refined cosmological parameter constraints derived from a cosmic shear analysis of the fourth data release of the Kilo-Degree Survey (KiDS-1000). Our main improvements include enhanced galaxy shape measurements made possible by an updated version of the *lens*fit code and improved shear calibration achieved with a newly developed suite of multi-band image simulations. Additionally, we incorporated recent advancements in cosmological inference from the joint Dark Energy Survey Year 3 and KiDS-1000 cosmic shear analysis. Assuming a spatially flat standard cosmological model, we constrain $S_8 \equiv \sigma_8 (\Omega_m/0.3)^{0.5} = 0.776^{+0.029+0.002}_{-0.027-0.003}$, where the second set of uncertainties accounts for the systematic uncertainties within the shear calibration. These systematic uncertainties stem from minor deviations from realism in the image simulations and the sensitivity of the shear measurement algorithm to the morphology of the galaxy sample. Despite these changes, our results align with previous KiDS studies and other weak lensing surveys, and we find a ~2.3σ level of tension with the *Planck* cosmic microwave background constraints on S_8.

Key words. cosmological parameters – cosmology: observations – gravitational lensing: weak – surveys

1. Introduction

Weak gravitational lensing by large-scale structure, also known as cosmic shear, is a powerful technique for studying the matter distribution in the Universe without assuming a specific correlation between dark and baryonic matter (e.g. Blandford et al. 1991; Miralda-Escude 1991; Kaiser 1992)[1]. Owing to its remarkable potential in exploring the cosmic matter distribution, cosmic shear analysis has gained popularity since its first detection over 20 years ago (Bacon et al. 2000; Kaiser et al. 2000; Van Waerbeke et al. 2000; Wittman et al. 2000). When distance information for source galaxies is also known, we can differentiate between them along the line of sight and perform a tomographic analysis, which entails reconstructing the 3D matter distribution from multiple 2D projections. This tomographic cosmic shear analysis is especially effective for constraining dark energy properties, as it sheds light on the evolution of cosmic structures (e.g. Hu 1999; Huterer 2002).

Recent surveys, such as the Kilo-Degree Survey (KiDS; de Jong et al. 2013), the Dark Energy Survey (DES; Dark Energy Survey Collaboration 2016), and the Hyper Suprime-Cam (HSC) survey (Aihara et al. 2018), primarily focus on constraining the amplitude of matter density fluctuations. Conventionally, this quantity is characterised by the parameter $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5}$, where Ω_m is the matter density parameter and σ_8 is the standard deviation of matter density fluctuations in spheres of radius $8\,h^{-1}$ Mpc, computed using linear theory, where the Hubble constant $H_0 = 100\,h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Interestingly, the S_8 values derived from these weak lensing surveys are consistently lower than those predicted by cosmic microwave background (CMB) observations from the *Planck* satellite.

Specifically, the latest cosmic shear analyses from KiDS ($0.759^{+0.024}_{-0.021}$; Asgari et al. 2021, hereafter A21), DES ($0.759^{+0.025}_{-0.023}$; Amon et al. 2022; Secco et al. 2022), and HSC ($0.769^{+0.031}_{-0.034}$; Li et al. 2023a; $0.776^{+0.032}_{-0.033}$; Dalal et al. 2023)

---

[1] However, with increasing precision in weak lensing observations, the impact of baryonic processes, such as radiative cooling and feedback from star formation and active galactic nuclei, on the observed matter distribution can no longer be ignored for small-scale structures (e.g. van Daalen et al. 2011; Semboloni et al. 2011).

provide S_8 values that are roughly 2σ lower than the *Planck* predictions (0.832 ± 0.013; Planck Collaboration VI 2020) based on the standard spatially flat Λ cold dark matter (ΛCDM) cosmological model. Most recently, a joint cosmic shear analysis of the DES Year 3 data and the fourth data release of KiDS by the two survey teams (DES and KiDS Collaboration 2023, hereafter DK23) yields an S_8 constraint of $0.790^{+0.018}_{-0.014}$, which is closer to the *Planck* results but still shows a level of 1.7σ difference. This mild difference in the S_8 constraints between the weak lensing surveys and CMB observations triggered extensive discussions from various perspectives, encompassing potential systematic errors in the data (e.g. Efstathiou & Lemos 2018; Köhlinger et al. 2019), the influence of the baryonic physics (e.g. Schneider et al. 2002; Amon & Efstathiou 2022; Preston et al. 2023), and a potential deviation from the standard ΛCDM model (see Perivolaropoulos & Skara 2022 for a recent review).

Here, we focus on the control of systematics in the cosmic shear analysis, particularly those arising during the KiDS shear measurement process. Measuring lensing-induced shear from noisy pixelised galaxy images is a challenging task, complicated further by distortions caused by the point spread function (PSF) resulting from instrumental and observational conditions, as well as blending effects that arise when two or more objects are close on the sky (see Mandelbaum 2018 for a review). These factors can introduce significant measurement biases (e.g. Paulin-Henriksson et al. 2008; Melchior & Viola 2012; Refregier et al. 2012; Massey et al. 2013; Dawson et al. 2016; Euclid Collaboration 2019) and alter the selection function of the source sample, leading to selection bias (e.g. Hartlap et al. 2011; Chang et al. 2013; Hoekstra et al. 2021). Therefore, obtaining unbiased shear measurements requires careful calibration, which can be performed using either pixel-level image simulations (e.g. Miller et al. 2013; Hoekstra et al. 2015; Fenech Conti et al. 2017, hereafter FC17; Samuroff et al. 2018; Mandelbaum et al. 2018) or the data themselves (e.g. Huff & Mandelbaum 2017; Sheldon & Huff 2017; Sheldon et al. 2020).

Additionally, in the case of large-area imaging surveys, determining the distance information for individual source galaxies depends on redshifts derived from broadband photometric observations. These photometric redshift estimates, which are subject to significant uncertainty, require careful calibration using spectroscopic reference samples (e.g. Hoyle et al. 2018; Tanaka et al. 2018; Hildebrandt et al. 2021). Furthermore, recent studies have shown that the blending of source images results in the coupling of shear and redshift biases (e.g. MacCrann et al. 2022; Li et al. 2023b, hereafter L23). Consequently, a joint calibration of these two estimates becomes essential, which will necessitate the use of multi-band image simulations in future cosmic shear analyses.

In light of all these concerns, we implemented several improvements to the cosmic shear measurements in KiDS, as detailed in L23. We enhanced the accuracy of the galaxy shape measurements by using an upgraded version of the *lens*fit code (Miller et al. 2007, 2013; Kitching et al. 2008), complemented by an empirical correction scheme that reduces PSF contamination. More notably, in L23 we introduced SKiLLS (SURFS-based KiDS-Legacy-Like Simulations), a suite of multi-band image simulations that enables a joint calibration of shear and redshift estimates. This is an important element for the forthcoming weak lensing analysis of the complete KiDS survey, known as the KiDS-Legacy analysis (Wright et al., in prep.).

In this paper we take an intermediate step towards the forthcoming KiDS-Legacy analysis by applying the improvements from L23 to a cosmic shear analysis based on the fourth data release of KiDS (Kuijken et al. 2019). In contrast to previous KiDS cosmic shear analyses, which used shear calibration methods developed in FC17 and Kannawadi et al. (2019, hereafter K19) based on single-band image simulations, our analysis adopted SKiLLS, marking the first instance of multi-band image simulations being used for KiDS cosmic shear analysis[2]. We also incorporated recent advancements in cosmological inference and updated the current cosmological parameter constraints from KiDS. In particular, we updated the code for the non-linear evolution of the matter power spectrum calculation from HMCODE to the latest HMCODE-2020 version (Mead et al. 2021). We also investigated the impact of the intrinsic alignment (IA) model by incorporating amplitude priors inspired by Fortuna et al. (2021a).

The remainder of this paper is structured as follows. In Sect. 2 we introduce and validate the updated KiDS shear catalogue, which is followed by the shear and redshift calibration in Sect. 3. We describe our cosmological inference method in Sect. 4 and present the results in Sect. 5. Finally, we summarise the results in Sect. 6.

2. Updated weak lensing shear catalogue

Our shear catalogue is based on the fourth data release of KiDS (Kuijken et al. 2019), which combines optical observations in the *ugri* bands from KiDS using the European Southern Observatory (ESO) VLT Survey Telescope (de Jong et al. 2013) and near-infrared observations in the $ZYJHK_s$ bands from the ESO Visible and Infrared Survey Telescope for Astronomy (VISTA) Kilo-degree INfrared Galaxy (VIKING) survey (Edge et al. 2013). The dataset covers 1006 deg^2 survey tiles and includes nine-band photometry measured using the Gaussian Aperture and PSF (GAAP) pipeline (Kuijken et al. 2015). The photometric redshifts (photo-zs) for individual source galaxies were estimated using the Bayesian photometric redshift (BPZ) code (Benítez 2000). After masking, the effective area of the dataset in the charge-coupled device (CCD) pixel frame is 777.4 deg^2 (Giblin et al. 2021). To perform the cosmic shear analysis, we divided the source sample into five tomographic bins based on the BPZ estimates (z_B). The first four bins have a spacing of $\Delta z_B = 0.2$ in the range $0.1 < z_B \le 0.9$, while the fifth bin covers the range $0.9 < z_B \le 1.2$, following the previous KiDS cosmic shear analyses.

2.1. Galaxy shapes measured with the updated lensfit

When preparing the shear measurements for the upcoming data release of KiDS, we upgraded the *lens*fit code (Miller et al. 2007, 2013; Kitching et al. 2008) from version $309c$ to version 321 (see L23 for details). The latest version includes a correction to an anisotropic error in the original likelihood sampler, which previously caused a small yet noticeable residual bias that was not related to the PSF or underlying shear (Miller et al. 2013; Hildebrandt et al. 2016; Giblin et al. 2021). We used the new code to re-measure the galaxy shapes, resulting in a new shear catalogue. Throughout the paper, we refer to the new shear catalogue as KiDS-1000-v2 to distinguish it from the previous KiDS-1000(-v1) shear catalogue (Giblin et al. 2021).

---

[2] K19 did attempt to assign photo-z estimates from data to simulations, but the actual photo-z measurements were not simulated.

The raw measurements from the *lens*fit code suffer from biases primarily due to the PSF anisotropy, but also because of the object selection and weighting scheme. To address these biases, FC17 introduced an empirical correction scheme to isotropise the original measurement weights, which was used in previous KiDS studies (see also K19). This correction scheme mitigates the *lens*fit weight biases and reduces the bias induced by the PSF anisotropy to an acceptable level. However, notable residual biases still persist (Giblin et al. 2021). Moreover, L23 find that the method is susceptible to variations in the sample size, posing challenges for consistent application to both data and simulations.

Therefore, a new correction scheme was introduced by L23 that modifies both the measured ellipticities and weights to ensure the average PSF leakage, defined as the fraction of the PSF ellipticity leaking into the shear estimator, is negligible in each tomographic bin. For further details, we direct readers to L23. In summary, the new correction scheme first isotropises the measurement weights, then adjusts the measured ellipticities to eliminate any remaining noise bias and selection effects. We note that this correction scheme is not designed to refine the shape measurements of individual galaxies; rather, it aims to ensure that the collectively weighted shear signal is robust against PSF leakage. In this work, we applied this newly developed empirical correction to the KiDS-1000-v2 shear catalogue.

## 2.2. Validation of the shear estimates

In order to use the weak lensing shear catalogue for cosmological inference, it is crucial to first verify the accuracy of the shear estimation and ensure that the residual contamination from systematic effects is within the acceptable level for scientific analysis. To achieve this, Giblin et al. (2021) proposed a series of null-tests to assess the robustness of the KiDS-1000-v1 shear catalogue. With the updated galaxy shape measurements in the KiDS-1000-v2 catalogue, it is necessary to repeat some of these tests to confirm the reliability of the new catalogue.

As the KiDS-1000-v2 catalogue updates only the galaxy shape measurements while maintaining the established photometry and PSF models, we did not repeat tests related to photometry and PSF modelling. We started by examining the PSF leakage in the weighted *lens*fit shear estimator, using the first-order systematics model proposed by Heymans et al. (2006). This model takes the form (Giblin et al. 2021)

$$\epsilon_k^{\mathrm{obs}} = (1 + m_k)(\epsilon_k^{\mathrm{int}} + \gamma_k) + \alpha_k \epsilon_k^{\mathrm{PSF}} + c_k, \quad [k = 1, 2], \quad (1)$$

where ϵ^{obs} denotes the measured galaxy ellipticity, m is the multiplicative shear bias[3], ϵ^{int} refers to the intrinsic galaxy ellipticity, γ stands for the cosmic shear signal (which is the parameter of interest), α is the PSF leakage factor, and c is an additive term comprising residual biases unrelated to the PSF or underlying shear. The subscript $k = 1, 2$ denotes the two ellipticity components. We note that we did not include PSF modelling errors in Eq. (1), as we used the same PSF model as Giblin et al. (2021), who had already confirmed its accuracy. Assuming that $(\epsilon_k^{\mathrm{int}} + \gamma_k)$ averages to zero for a large galaxy sample (a property validated with the KiDS data; see, for example, Sect. 3 in Giblin et al. 2021), we can determine the α and c parameters from the data using a simple linear regression method.
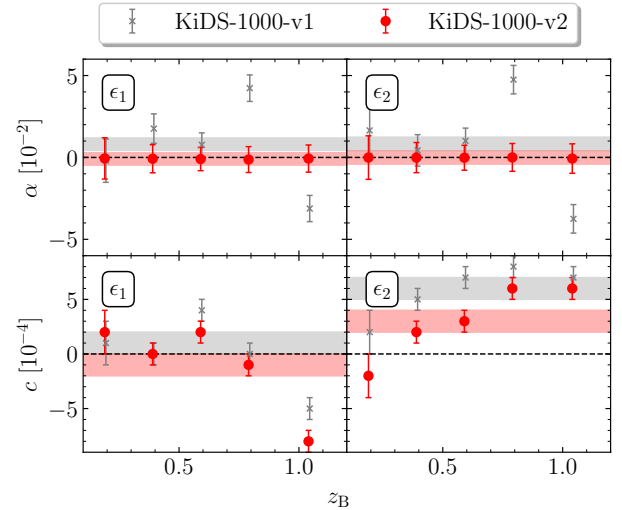
Fig. 1. PSF contamination, α (top panels) and additive term, c (bottom panels) as a function of tomographic bin labelled with the central z_{B} value. The measurements are obtained from a weighted linear fitting using Eq. (1). The red points represent measurements from the KiDS-1000-v2 catalogue, while the grey points show the measurements from the KiDS-1000-v1 catalogue. The red and grey bars correspond to results from the entire sample without tomographic binning.

Figure 1 presents the measured PSF leakage α and the additive term c for the KiDS-1000-v2 catalogue, alongside the measurements from the KiDS-1000-v1 catalogues for comparison. As expected, the KiDS-1000-v2 catalogue exhibits a mean α-term consistent with zero for all redshift bins, owing to the empirical correction scheme outlined in Sect. 2.1 (see also Sect. 4 in L23). The upgraded *lens*fit code has reduced the overall c_2-term by half, reaching a level of $c_2 \sim (3 \pm 1) \times 10^{-4}$ for the entire sample. However, despite this improvement, the c term has not been eliminated, particularly in distant tomographic bins where a small but noticeable c term still persists, which was not seen in the simulations.

To correct for these residual small additive c-terms, we used the same empirical correction method as in previous KiDS analyses. Specifically, we subtracted the weighted average ellipticity from the observed ellipticity for each redshift bin as $\epsilon_{\mathrm{corr}}^{\mathrm{obs}} = \epsilon^{\mathrm{obs}} - \overline{\epsilon^{\mathrm{obs}}}$. Nevertheless, we caution that subtracting the mean c-term does not guarantee the removal of all additive biases, especially when detector-level effects, such as "charge transfer inefficiency" (e.g. Rhodes et al. 2007; Massey 2010) and "pixel bounce" (e.g. Toyozumi & Ashley 2005), can introduce position-dependent bias patterns. Although we have detected such effects in KiDS data (Hildebrandt et al. 2020; Giblin et al. 2021), their level does not affect the current cosmic shear analysis. More specifically, Asgari et al. (2019) show that even if current detector-level effects were increased by a factor of 10, they would not cause significant bias for KiDS-like analyses.

The cosmic shear signal is conventionally measured using the two-point shear correlation function, defined as[4]

$$\hat{\xi}_{\pm}^{ij}(\theta) = \frac{\sum_{ab} w_a w_b \left[\epsilon_t^i(\boldsymbol{x}_a) \epsilon_t^j(\boldsymbol{y}_b) \pm \epsilon_\times^i(\boldsymbol{x}_a) \epsilon_\times^j(\boldsymbol{y}_b) \right]}{\sum_{ab} w_a w_b}, \quad (2)$$

---

[3] Throughout this paper, we interchangeably use "multiplicative bias" and "shear bias", as our simulation-based shear calibration only addresses this parameter. Conversely, PSF leakage and the additive term are empirically corrected.

[4] In this study, all measurements of the two-point shear correlation function are conducted using the TREECORR code (Jarvis et al. 2004; Jarvis 2015).
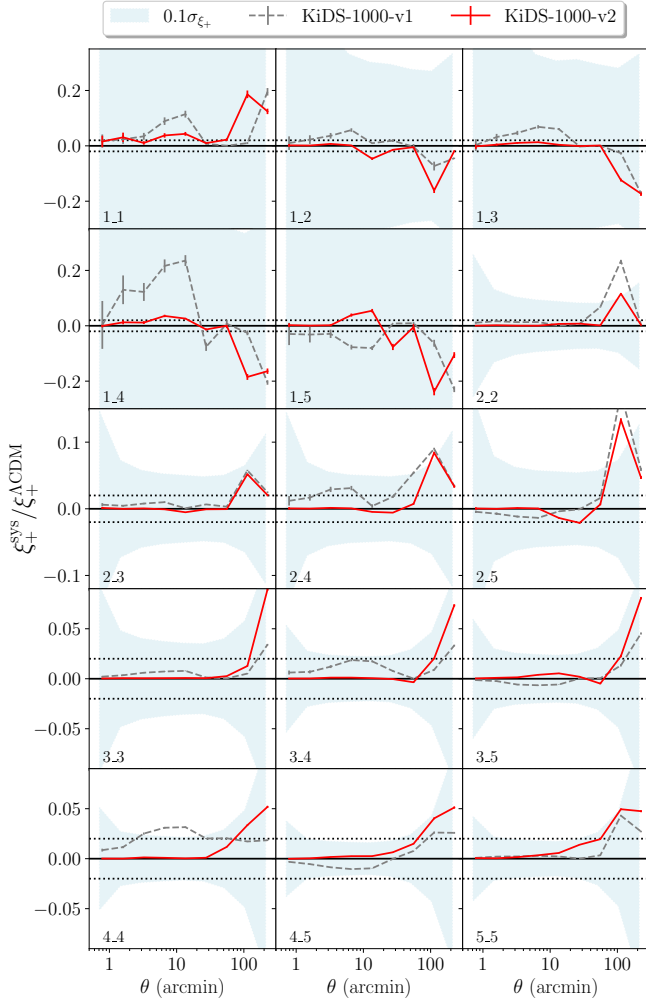
Fig. 2. Ratio of the PSF contamination, $\xi_+^{\rm sys}$, computed using Eq. (3) to the predicted amplitude of the cosmic shear signal, $\xi_+^{\Lambda\rm CDM}$, across all 15 tomographic bin combinations. The red lines depict results from the KiDS-1000-v2 catalogue, whereas the grey lines show those from the KiDS-1000-v1 catalogue. The blue shaded regions represent a range of ±10% of the standard deviation of the measured cosmic shear signal. This deviation is determined from the covariance matrix using statistics from the KiDS-1000-v2 catalogue. The dotted horizontal lines indicate the 2% level of the predicted cosmic shear signal.

where θ represents the separation angle between a pair of galaxies (a, b), the tangential and cross ellipticities $\epsilon_{t,\times}$ are computed with respect to the vector $\boldsymbol{x}_a - \boldsymbol{y}_b$ that connects the galaxy pair, and the associated measurement weight is denoted by w. Therefore, it is crucial to examine the systematics in the two-point statistics. Following the method of Bacon et al. (2003), we estimated the PSF leakage into the two-point correlation function measurement using

$$\xi_\pm^{\rm sys} = \frac{\langle\epsilon^{\rm obs}\epsilon^{\rm PSF}\rangle^2}{\langle\epsilon^{\rm PSF}\epsilon^{\rm PSF}\rangle}, \qquad (3)$$

where the $\langle\cdot\rangle$ represents the correlation function.

In Fig. 2 we present the ratio of the measured $\xi_+^{\rm sys}$ to the theoretical predictions of the cosmic shear signal. The blue shaded region denotes ±10% of the standard deviation of the cosmic shear signal, extracted from the analytical covariance. This covariance is calculated using an independent implementation of the methodology of Joachimi et al. (2021), and it incor-
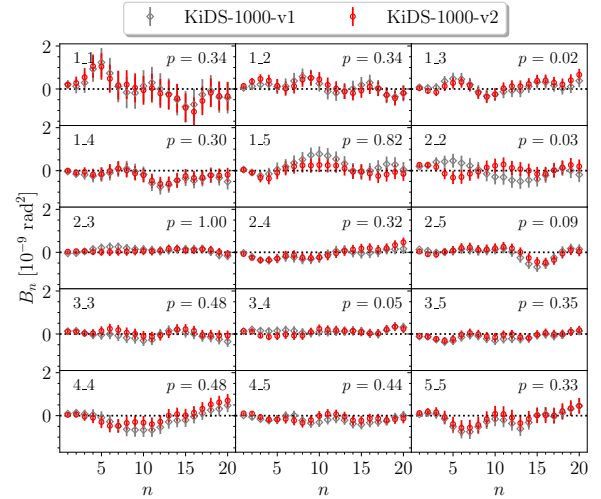
Fig. 3. Measurements of the B-mode signals using COSEBIs for the KiDS-1000-v2 catalogue (red points) compared to the KiDS-1000-v1 catalogue (grey points). The error bars originate from the diagonal of an analytical covariance matrix, accounting solely for measurement noise. For the KiDS-1000-v2 catalogue, we re-calculated the covariance using the method introduced by Joachimi et al. (2021), incorporating the updated statistics. The p-values for the KiDS-1000-v2 catalogue, shown in the top-right corner of each panel, were calculated with 20 degrees of freedom, which corresponds to the number of modes used in each correlation.

porates the sample statistics of the updated catalogue. We compared the results from the KiDS-1000-v2 catalogue with those from the KiDS-1000-v1 catalogue. We observe general improvements, particularly in the high-redshift bins, where the PSF contamination is now negligible. The only exceptions are found in some large-scale bins ($\theta > 60\,\rm arcmin$), where the expected fiducial cosmic shear signal is relatively small and overwhelmed by high statistical noise.

To the leading order, the weak lensing effect introduces only curl-free gradient distortions (E-mode signal), which makes the curl distortions (B-mode signal) a useful null-test for residual systematics in the shear measurement[5]. Following the convention of KiDS (Hildebrandt et al. 2017; Giblin et al. 2021), we used the complete orthogonal sets of E/B-integrals (COSEBIs; Schneider et al. 2010) to measure the B-mode signal. The COSEBIs provide an optimal E/B separation by combining different angular scales from the $\hat{\xi}_\pm$ measurements.

Figure 3 presents the measured B-mode signals for all combinations of tomographic bins in our analysis, alongside the B-mode measurements from the KiDS-1000-v1 catalogue for comparison. To enable a direct comparison, we used the same scale range of $(0\!.\!5, 300')$ as in Giblin et al. (2021) for calculating the COSEBIs B-mode[6]. Assuming a null signal, we computed the p-value for each B-mode measurement, setting the degrees of freedom equal to the number of modes in each measurement ($n = 20$). The covariance matrix, accounting only for shot noise,

---

[5] Some higher-order effects from lensing, such as source redshift clustering (e.g. Schneider et al. 2002), and IA of nearby galaxies (e.g. Troxel & Ishak 2015; Joachimi et al. 2015) can also introduce B-mode signals. However, these contributions are expected to be negligible for current weak lensing surveys (e.g. Hilbert et al. 2009).

[6] We also evaluated an alternate scale range of $(2', 300')$, consistent with our fiducial cosmic shear analysis. As anticipated, the B-mode signal was more negligible in this scenario due to reduced small-scale contamination.

was estimated using an analytical model from Joachimi et al. (2021) applied to the updated catalogue. It is noteworthy that our covariance matrix differs from the one used in Giblin et al. (2021). This is due to the changes in sample statistics resulting from the updated shape measurement code and redshift calibration relative to the KiDS-1000-v1 catalogue used in Giblin et al. (2021). Most diagonal entries in our matrix show reduced uncertainties, ranging from a level of percent to ten percent. Therefore, if the absolute systematic levels are comparable between the two catalogues, our test would likely show a slight increase in the final p-values compared to those in Giblin et al. (2021). As indicated in the top-right corner of each panel, the estimated p-values suggest that the measured B-mode signals align with a null signal across all bin combinations. The lowest p-value, $p = 0.02$, was found in the cross-correlation between the first and third tomographic bins.

After conducting all these tests, we can conclude that the KiDS-1000-v2 catalogue has reduced systematics when compared to the results from the KiDS-1000-v1 catalogue. These improvements are largely attributed to the updated version of the *lens*fit code, as well as the implementation of a new empirical correction scheme for PSF contamination. These results give us the confidence to use the updated catalogue for cosmological inference.

3. Shear and redshift calibration

The main improvement in our calibration comes from the use of SKiLLS multi-band image simulations, as developed in L23. These simulations fuse cosmological simulations with high-quality observational data to create mock galaxies with photometric and morphological properties closely resembling real-world galaxies. The observational data used by SKiLLS, drawn from the catalogue of Griffith et al. (2012), is identical to that used in K19. In L23, we developed a vine-copula-based algorithm that learns the measured morphological parameters from this catalogue and assigns them to the SURFS-Shark mock galaxies (Elahi et al. 2018; Lagos et al. 2018). We verified that the learning procedure maintains the observed multi-dimensional correlations between morphological parameters, magnitude, and redshifts. Nevertheless, both the observed catalogue from Griffith et al. (2012) and the learning algorithm possess inherent limitations, resulting in unavoidable uncertainties in our simulation input catalogue. These uncertainties are addressed in our shear calibration in Sect. 3.2.

To create KiDS+VIKING-like nine-band images, SKiLLS replicated the instrumental and observational conditions of 108 representative tiles selected from six sky pointings evenly distributed across the footprint of the KiDS fourth data release. The star catalogue was generated for each sky pointing using the TRILEGAL population synthesis code (Girardi et al. 2005) to account for the variation in stellar densities across the footprint. For the primary r-band images, on which the galaxy shapes were measured, SKiLLS included the correlated pixel noise introduced by the stacking process and the PSF variation between CCD images.

On the data processing side, SKiLLS followed the entire KiDS procedure, including object detection, PSF homogenisation, forced multi-band photometry, photo-z estimation, and shape measurements. The end result is a self-consistent joint shear-redshift mock catalogue that matches KiDS observations in both shear and redshift estimates. By taking this end-to-end approach, we accounted for photo-z-related selection effects in our shear bias estimation and enabled redshift calibration using

the same mock catalogue. While our current analysis focuses on the improvement in shear calibration, it represents an intermediate step towards the KiDS-Legacy analysis, which will implement joint shear and redshift calibrations facilitated by the SKiLLS mock catalogue.

3.1. Calibration

To correct for shear bias in our measurements, we adopted the method used in previous KiDS studies (FC17, K19). For each tomographic bin i, we applied an average shear bias correction factor, $\overline{m^i}$, which is derived by averaging the individual m values of all sources within the respective tomographic bin. These individual m values are determined using Eq. (1), based on simulations mapped on a grid of the *lens*fit reported model signal-to-noise ratio and resolution. Here, the resolution is defined as the ratio of the PSF size to the measured galaxy size. To align our simulations more closely with the target data, we followed KiDS conventions and re-weighted the simulation estimates according to the grid of signal-to-noise ratio and resolution. Further details about the re-weighting procedure can be found in Sect. 5.1 of L23.

Although the averaging method addresses the noise in individual source's m estimation, it does not account for correlations involving shear bias. Thus, we have $\langle[1+m^i(\theta')][1+m^j(\theta'+\theta)]\rangle = (1 + \overline{m^i})(1 + \overline{m^j})$, with θ and θ' representing different separation angles between galaxy pairs. To test this assumption, we directly measured $\langle[1 + m^i(\theta')][1 + m^j(\theta' + \theta)]\rangle$ from image simulations and compared it to $(1+\overline{m^i})(1+\overline{m^j})$. Further details on this test can be found in Appendix A. In summary, we find a negligible difference between the two estimators, a result that falls well within the current KiDS requirements. This validates the assumption for the KiDS analysis.

Given that the updated galaxy shape measurements also lead to changes in the sample selection function, it is necessary to repeat the redshift calibration for the KiDS-1000-v2 catalogue, even though our primary focus is to improve shear calibration. To quantify the changes in galaxy samples introduced by the modifications in shape measurements from the KiDS-1000-v1 to KiDS-1000-v2 catalogues, we compared their effective number densities before applying any redshift calibration. The observed percentage differences in each tomographic bin, from low to high redshift bins, are -1.8%, -0.4%, 0.2%, 1.3%, and 3.2%. Here, negative values indicate a decrease in density from the v1 to the v2 catalogue, while positive values signify an increase. These differences are largely attributed to changes in the weighting scheme brought by the *lens*fit updates, as well as the implementation of the new empirical correction scheme for PSF leakage, as discussed in Sect. 3 and in L23. For this, we employed a methodology identical to the one used by Wright et al. (2020), Hildebrandt et al. (2021), and van den Busch et al. (2022, hereafter vdB22). It is based on a direct calibration method (Lima et al. 2008) implemented with a self-organising map (SOM; Kohonen 1982; Masters et al. 2015). More information on our implementation is provided in Appendix B, while Wright et al. (2020), Hildebrandt et al. (2021), and vdB22 offer more comprehensive discussions.

The SOM-based redshift calibration method uses a "gold selection" criterion to filter out sources that are not represented in the spectroscopic reference sample (see Appendix B). However, this process influences shear biases as it alters the selection function of the final sample. To ensure a consistent estimation of shear biases, we created the SKiLLS-gold catalogue by mimicking this quality control on the SKiLLS mock catalogue, using

Table 1. Data properties for the KiDS-1000-v2 catalogue.

Bin	Photo-z range	n_{eff} [arcmin^{-2}]	$\sigma_{\epsilon,i}$	$\delta z = z_{\mathrm{est}} - z_{\mathrm{true}}$	m_{raw}	m_{final}	σ_{m}
1	$0.1 < z_{\mathrm{B}} \leq 0.3$	0.68	0.27	0.000 ± 0.0096	-0.023	-0.021	0.019
2	$0.3 < z_{\mathrm{B}} \leq 0.5$	1.30	0.26	0.002 ± 0.0114	-0.025	-0.023	0.008
3	$0.5 < z_{\mathrm{B}} \leq 0.7$	1.97	0.28	0.013 ± 0.0116	-0.013	-0.015	0.007
4	$0.7 < z_{\mathrm{B}} \leq 0.9$	1.39	0.27	0.011 ± 0.0084	0.018	0.015	0.006
5	$0.9 < z_{\mathrm{B}} \leq 1.2$	1.35	0.29	-0.006 ± 0.0097	0.032	0.031	0.006

Notes. Comparable summary statistics for the KiDS-1000-v1 catalogue can be found in Table 1 of A21. We note that the differences in summary statistics between our work and A21 stem from both the updated *lens*fit code and the enhanced redshift calibration outlined in vdB22. The effective number density n_{eff} and the ellipticity dispersion per ellipticity component $\sigma_{\epsilon,i}$ are calculated using the formulae provided in Appendix C of Joachimi et al. (2021). The n_{eff} values in this table are derived from an effective area of 777.4 sq. deg. in the CCD pixel frame, making them directly comparable to the values in Table 1 of A21. The correlated Gaussian redshift priors are based on the differences between the estimated and true redshifts, $\delta z = z_{\mathrm{est}} - z_{\mathrm{true}}$, as reported in vdB22. The priors are denoted as $\mu_i \pm \sigma_i$, where μ_i represents the mean shift and σ_i corresponds to the square root of the covariance matrix's diagonal elements. The m_{raw} results are derived from idealised constant shear simulations, while the m_{final} results, our fiducial outcomes, include corrections for the shear-interplay effect and PSF modelling bias. Statistical uncertainties, determined by the simulation volume, are directly computed from the fiducial simulations and denoted as σ_{m}.

the same SOM trained by the spectroscopic reference sample as the real data. We derived the appropriate shear bias correction factors from this SKiLLS-gold catalogue for individual tomographic bins, and present these values in Table 1. It is worth noting that the shear bias estimates presented in this work differ slightly from those in L23, which did not include the gold selection procedure. Despite this, the differences in the estimated shear biases are relatively minor across all tomographic bins, with the first tomographic bin showing the most noticeable change of 0.008.

Our fiducial results, m_{final}, account for the impact of PSF modelling uncertainties and the "shear interplay" effect, which occurs when galaxies from different redshifts are blended together. For more details on these effects, we refer the reader to L23 and MacCrann et al. (2022). Additionally, we provide the idealised m_{raw} results, which do not consider these higher-order effects. By comparing the cosmological constraints obtained from these two cases, we aim to evaluate the robustness of previous KiDS results with respect to these higher-order effects, which were not taken into account in the earlier shear calibration (FC17; K19).

3.2. Calibration uncertainties

Systematic uncertainties arising from redshift and shear calibrations can propagate into cosmological analyses, potentially leading to biased results. Therefore, it is crucial to adequately address these uncertainties in the analysis. In this section we outline our approach to managing these calibration uncertainties.

The uncertainties in redshift calibration were addressed by introducing an offset parameter for the estimated mean redshift of galaxies in each tomographic bin. This offset parameter, described as correlated Gaussian priors, serves as a first-order correction to both the statistical and systematic uncertainties associated with redshift calibration. Table 1 lists the exact values for these parameters, which we obtained from vdB22 and Hildebrandt et al. (2021). They determined these prior values using spectroscopic and KiDS-like mock data generated by van den Busch et al. (2020). We consider the current priors to be conservative enough to account for any potential changes in the redshift biases from KiDS-1000-v1 to KiDS-1000-v2, given that both catalogues use the same photometric estimates. However, for the forthcoming KiDS-Legacy analysis, we plan to re-estimate these values based on the new SKiLLS mock data.

We improved our approach to handling uncertainties related to the shear calibration. In L23, nominal uncertainties were proposed for each tomographic bin based on sensitivity analyses. This aimed to ensure the robustness of the shear calibration within the specified uncertainties, but at the cost of reducing statistical power. In this work, we aim to improve this approach by separately accounting for the statistical and systematic uncertainties within the shear calibration.

The statistical uncertainties, as presented in Table 1, are computed directly from simulations and are limited only by the volume of the simulations, which can be increased with more computing resources[7]. These uncertainties are also easily propagated into the covariance matrix for cosmological inference. Although increasing the simulation volume could, in principle, reduce these uncertainties, we find that the current values already comfortably meet the KiDS requirements; thus, further efforts in this direction were considered unnecessary.

If the SKiLLS simulations perfectly match KiDS data, these statistical uncertainties would be the only contribution to the final uncertainty from the shear calibration. However, since our simulations are not a perfect replica of the real observations, residual shear biases may still be present in the data even after calibration. These biases, referred to as systematic uncertainties, are typically the primary source of error in shear calibration. Increasing the simulation volume cannot improve these uncertainties as they are determined by the realism of the image simulations. The level of these uncertainties can only be roughly estimated through sensitivity analyses.

Since the systematic residual shear biases directly scale the data vector, accurately quantifying their impact using the covariance matrix is challenging. Therefore, we used a forward modelling approach to capture the impact of these systematic uncertainties. Instead of incorporating these uncertainties into the covariance matrix, we examined how the final estimates of the cosmological parameters change due to the shift in signals caused by the systematic residual shear biases. This forward modelling approach can be easily implemented using simple optimisation algorithms since the shift is small, and the covariance remains unchanged. More details on how to determine residual shear biases and implement the forward modelling approach are provided in Appendix C.

---

[7] However, the finite volume of the input galaxy sample prevents an indefinite increase.

4. Cosmological inference

The cosmological inference in this study largely aligns with the approach used in the KiDS-1000-v1 analyses (A21; vdB22), with minor modifications primarily influenced by the recent joint DES Y3+KiDS-1000 cosmic shear analysis (DK23). In this section, we outline the configurations and reasoning behind these choices in our fiducial analysis. For certain notable changes, we also conducted extended analysis runs with different configurations to evaluate the impact of these modifications. Our analysis code is publicly accessible[8].

Our code builds upon the `Cat_to_Obs_K1000_P1`[9] and the KiDS Cosmology Analysis Pipeline (KCAP)[10] infrastructure, as developed in Giblin et al. (2021), Joachimi et al. (2021), A21, Heymans et al. (2021), and Tröster et al. (2021). The `Cat_to_Obs_K1000_P1` pipeline converts KiDS shear and redshift measurements into various second-order statistics, with the assistance of the TREECORR code (Jarvis et al. 2004; Jarvis 2015). Meanwhile, KCAP estimates cosmological parameters using the COSMOSIS framework, which bridges the likelihood function calculation pipelines with sampling codes (Zuntz et al. 2015).

We measured the shear field using COSEBIs (Schneider et al. 2010). As reported by Asgari et al. (2020), COSEBIs offer enhanced robustness against small-scale effects on the shear power spectrum, which primarily stem from complex baryon feedback. Furthermore, we accounted for baryon feedback when modelling the matter-matter power spectrum using HMCODE-2020 (Mead et al. 2021) within the CAMB framework with the version 1.4.0 (Lewis et al. 2000; Howlett et al. 2012).

HMCODE-2020, an updated version of HMCODE (Mead et al. 2015, 2016), models the non-linear matter-matter power spectrum, incorporating the influence of baryon feedback through an enhanced halo-model formalism. This updated model is empirically calibrated using hydrodynamical simulations, following a more physically informed approach. Unlike its predecessor calibrated with OWLS hydrodynamical simulations (van Daalen et al. 2011), this newer version uses the updated BAHAMAS hydrodynamical simulations for calibration (McCarthy et al. 2017). These simulations, in turn, are calibrated to reproduce the observed galaxy stellar mass function and the hot gas mass fractions of groups and clusters. This calibration ensures that the simulation accurately reflects the impact of feedback on the overall distribution of matter (refer to McCarthy et al. 2017 for further details). Furthermore, HMCODE-2020 improves the modelling of baryon-acoustic oscillation damping and massive neutrino treatment, achieving an improved accuracy of 2.5% (compared to the previous version's 5%) for scales $k < 10\,h\,\mathrm{Mpc}^{-1}$ and redshifts $z < 2$ (Mead et al. 2021).

The model incorporates a single-parameter variant, T_{AGN}, representing the heating temperature of active galactic nuclei (AGNs). Higher T_{AGN} values correspond to more intense AGN feedback, leading to a lower observed matter power spectrum. Following DK23, we used a uniform prior on $\log_{10}(T_{\mathrm{AGN}})$ that ranged from 7.3 to 8.0. This choice was motivated by the findings from the BAHAMAS hydrodynamical simulations (McCarthy et al. 2017; van Daalen et al. 2020).

Given the characteristics of COSEBIs and the implementation of the HMCODE, the KiDS-1000-v1 analyses included small-scale measurements down to $\theta_{\min} = 0\farcs5$. This strategy was, how-

ever, re-evaluated in DK23, who suggest more stringent scale cuts for the KiDS COSEBIs data vector, determined by the baryon feedback mitigation strategy proposed by Krause et al. (2021). Following this recommendation, we applied a scale cut of $\theta_{\min} = 2'$ in our fiducial analysis.

We used the non-linear linear alignment (NLA) model to describe the IA of galaxies. This model combines the linear alignment model with a non-linear power spectrum and contains a single free parameter A_{IA} to describe the amplitude of IA signals (Hirata & Seljak 2004; Bridle & King 2007). It is also common to include a power law, with an index denoted as η_{IA}, to capture potential redshift evolution of the IA strength. To distinguish it from the redshift-independent NLA model, we refer to this variant as the NLA-z model.

In line with previous KiDS analyses, we took the redshift-independent NLA model as our fiducial choice since introducing η_{IA} has a minimal effect on the primary S_8 constraint (A21) and since current direct observations of IA signals show little evidence of substantial redshift evolution (e.g. Joachimi et al. 2011; Singh et al. 2015; Johnston et al. 2019; Fortuna et al. 2021b; Samuroff et al. 2023). However, Fortuna et al. (2021a) suggest that the selection of galaxy samples resulting from the redshift binning may introduce a detectable redshift variation in the IA signal, although its impact remains negligible for current weak lensing analyses. To assess the impact of η_{IA} on our results, we performed an extended run using the NLA-z model, following the same prior selection as in DK23.

The KiDS-1000-v1 analyses adopted a broad and uninformative prior for A_{IA}, ranging from $[-6, 6]$, considering that the data can constrain it and that an incorrect informative prior could bias the final cosmological results. Although uncertainties regarding IA signals remain large, recent developments in the field have improved our knowledge of the expected IA signal strength. For instance, Fortuna et al. (2021a) used a halo model formalism, incorporating results from the latest direct IA measurements, and predicted $A_{\mathrm{IA}} = 0.44 \pm 0.13$ for the redshift-independent NLA model targeted for KiDS-like mixed-colour lensing samples[11]. This prediction aligns well with the constraints from recent cosmic shear analyses (A21; Secco et al. 2022; Li et al. 2023a; Dalal et al. 2023). Moreover, recent studies revealed that other nuisance parameters in such analyses, especially those related to redshift calibration uncertainties, can result in misleading A_{IA} values (Hikage et al. 2019; Wright et al. 2020; Li et al. 2021; Fischbacher et al. 2023).

Given these considerations, we consider it necessary to explore the prior for the A_{IA} parameter. As an initial step towards a fully informed A_{IA} approach, we began by simply narrowing the previously broad prior, leaving a more comprehensive exploration of the IA model setups for the forthcoming KiDS-Legacy analysis. In our fiducial analysis, we chose a flat yet narrower prior of $[-0.2, 1.1]$, which corresponds to the 5σ credible region of predictions by Fortuna et al. (2021a). We note that our new prior will not significantly impact the sampling results, provided that the final posterior distributions fall within the set prior range. For comparison purposes, we also conducted a test run using the wider $[-6, 6]$ prior.

Sampling the high-dimensional posterior distribution is a challenging task. In the KiDS-1000-v1 analyses, an ellipsoidal nested sampling algorithm, MULTINEST (Feroz et al. 2009), was

---

[11] Fortuna et al. (2021a) also examined the NLA-z model under similar conditions, but found the fits were predominantly driven by the low-redshift bins, resulting in less accurate recovery of large-scale alignments at high redshifts.

used. However, recent studies demonstrated that MULTINEST systematically underestimates the 68% credible intervals for S_8 by about 10% in current weak lensing analyses (Lemos et al. 2023; DK23; Li et al. 2023a). A promising alternative is the sliced nested sampling algorithm, POLYCHORD (Handley et al. 2015a,b). It provides more accurate estimates of parameter uncertainties, making it our choice for the main analysis. However, it is worth noting that POLYCHORD is nearly five times slower than MULTINEST. Consequently, we retained MULTINEST for testing purposes. For our sampler settings, we followed DK23, adopting parameters $n_{live} = 500$, $n_{repeats} = 60$ and tolerance = 0.01 for POLYCHORD; and $n_{live} = 1000$, efficiency = 0.3, tolerance = 0.01, and constant efficiency = False for MULTINEST.

When presenting point estimates and associated uncertainties for parameter constraints, we adhere to the recommendations of Joachimi et al. (2021). We derived our best-fit point estimates from the parameter values at the maximum a posterior (MAP). Given that the MAP reported by the sampling code can be affected by noise due to the finite number of samples, we enhanced the precision of the MAP by conducting an additional local optimisation step. This process initiates from the MAP reported by the sampling code and utilises the Nelder–Mead minimisation method (Nelder & Mead 1965), a method also employed by A21. To represent uncertainties linked to these estimates, we computed the 68% credible interval based on the projected joint highest posterior density (PJ-HPD) region. This hybrid approach is more robust against projection effects stemming from high-dimensional asymmetric posterior distributions than traditional 1D marginal summary statistics (refer to Sect. 6 in Joachimi et al. 2021 for a comprehensive discussion). To facilitate comparison with results from other surveys, we also provide constraints based on the traditional mean and maximum of the 1D marginal posterior, along with their respective 68% credible intervals.

It is worth noting that, as systematic uncertainties from shear calibration are excluded in the construction of our covariance matrix (see Sect. 3.2), the uncertainties derived from the main sampling chains do not fully account for the true uncertainties. To compensate for the additional uncertainties arising from residual shear biases, we employed a forward modelling approach. This method involves shifting the data vector and subsequently the likelihood, based on the estimated residual shear biases, followed by recalculating the MAP. As the adjustment is minor and the covariance matrix remains static, it is not necessary to re-sample the posterior distribution. Instead, we simply needed to repeat the previously mentioned local optimisation step. Starting with the original MAP and using the updated likelihood, we can determine the new MAP corresponding to each shift in the data vector. The variation in these MAP estimates represents additional uncertainties introduced by the systematic uncertainties arising from shear calibration. Further details on this process can be found in Appendix C.

Table 2 summarises the model parameters and their priors as used in our fiducial analysis. These parameters can be broadly classified into two categories: the first category includes five cosmological parameters, which describe the spatially flat ΛCDM model we employed. We fixed the sum of the neutrino masses to a value of $0.06 \, \text{eV} \, c^{-2}$, where c is the speed of light. This choice is based on the Hildebrandt et al. (2020) finding of the negligible influence of neutrinos on cosmic shear analyses. The second category encompasses three nuisance parameters, accounting for astrophysical and measurement uncertainties as previously discussed. We note that all parameters, with the exception of T_{AGN}

Table 2. Fiducial model parameters and their priors.

Parameter	Symbol	Prior
Density fluctuation amp.	S_8	[0.1, 1.3]
CDM density	ω_c	[0.051, 0.255]
Baryon density	ω_b	[0.019, 0.026]
Hubble constant	h	[0.64, 0.82]
Scalar spectral index	n_s	[0.84, 1.1]
AGN heating temperature	$\log_{10}(T_{AGN}[\text{K}])$	[7.3, 8.0]
Intrinsic alignment amp.	A_{IA}	[−0.2, 1.1]
Redshift offsets	δ_z	$\mathcal{N}(\boldsymbol{\mu}; \boldsymbol{\sigma}^2)$

Notes. The first section lists the primary cosmological parameters describing the ΛCDM model assumed, while the second section contains nuisance parameters related to baryon feedback, intrinsic alignments, and redshift biases. The values in square brackets indicate the limits of top-hat priors. The notation $\mathcal{N}(\boldsymbol{\mu}; \boldsymbol{\sigma}^2)$ refers to a normal prior with mean $\boldsymbol{\mu}$ and (co-)variance $\boldsymbol{\sigma}^2$, as specified in Table 1.

and A_{IA}, retain the same priors as those used in the KiDS-1000-v1 cosmic shear analyses. The T_{AGN} parameter replaces the previous baryon feedback amplitude parameter associated with the preceding version of HMCODE, while the A_{IA} parameter adopts a narrower prior for reasons previously discussed.

5. Results

In this section we present our cosmological parameter constraints and evaluate the robustness of our findings against a variety of systematic uncertainties. We begin by presenting the outcomes from our fiducial analysis in Sect. 5.1. We then assess the impact of shear biases in Sect. 5.2, by quantifying the shifts in final constraints resulting from different shear bias scenarios. This highlights the main development of our work. Additionally, since we implemented several changes to the cosmological inference pipeline, we evaluate the effects of these adjustments by comparing results from multiple setup variations in Sect. 5.3.

5.1. Fiducial analysis results

Our fiducial model has a total of twelve free parameters: five are cosmological parameters specifying the spatially flat ΛCDM model with a fixed total neutrino mass, and the remaining seven are nuisance parameters addressing astrophysical and redshift calibration uncertainties, as detailed in Sect. 4. However, not all of these parameter are constrained by the cosmic shear analysis. In this section, we focus on the primary parameters that our analysis constrains. Meanwhile, the posterior distributions for all free parameters are displayed as contour plots in Appendix D for reference.

Table 3 provides the point estimates along with their corresponding 68% credible intervals for the primary parameter as constrained by our fiducial analysis using the POLYCHORD sampling code. We display results using three summary statistics: MAP and PJ-HPD, the mean of the 1D marginal posterior, and the maximum of the 1D marginal. As discussed in DK23, each of these approaches has its own advantages and limitations. Specifically, the accurate determination of MAP and PJ-HPD can be challenging, while marginal constraints for multi-dimensional posteriors are prone to projection effects. Aligning with the KiDS convention, we chose the MAP and PJ-HPD constraints as our headline results, but caution against direct comparisons

Table 3. Primary parameter constraints from our fiducial analysis, based on the KiDS-1000-v2 catalogue, as determined using the PolyChord sampling code.

Parameter	MAP and PJ-HPD	Marginal	
		Mean	Maximum
S_8	$0.776^{+0.029+0.002}_{-0.027-0.003}$	$0.765^{+0.029}_{-0.023}$	$0.769^{+0.027}_{-0.029}$
Ω_m	$0.259^{+0.115+0.001}_{-0.064-0.001}$	$0.302^{+0.062}_{-0.115}$	$0.273^{+0.102}_{-0.088}$
σ_8	$0.835^{+0.151+0.002}_{-0.158-0.003}$	$0.791^{+0.124}_{-0.163}$	$0.752^{+0.173}_{-0.129}$
A_{IA}	$0.348^{+0.350+0.009}_{-0.322-0.011}$	$0.400^{+0.330}_{-0.339}$	$0.397^{+0.340}_{-0.346}$

Notes. Our headline results, based on the MAP and PJ-HPD statistics, include additional uncertainties that account for systematic uncertainties within the shear calibration. These uncertainties, originating from minor deviations from realism in the image simulations and the shear measurement algorithm's sensitivity to the morphology of the galaxy sample, are estimated using a forward modelling approach (as detailed in Sect. 5.2). On the other hand, the statistical uncertainties within the shear calibration, determined by the simulation volume, are folded into the main uncertainties through their inclusion in the covariance matrix used for the cosmological inference. The mean-marginal is determined through `postprocess` within CosMoSIS using the default settings (Zuntz et al. 2015); while the max-marginal is calculated using the `ChainConsumer` with the settings of statistics = "max" and kde = 1.0 (Hinton 2016). The indicated uncertainties correspond to the 68% credible intervals.

with results from other surveys that might use different summary statistics. The uncertainties we report include additional contributions from the systematic uncertainties associated with our shear calibration, as detailed in Sect. 5.2. These additional uncertainties are overall small compared to the main sampling uncertainties, so when plotting the posterior distributions or conducting extended runs for test purposes, we did not incorporate these uncertainties.

Figure 4 shows the projected 2D posterior distributions for the parameters Ω_m and S_8, as derived from our fiducial setups employing PolyChord and MultiNest. We see that MultiNest results yield a roughly 10% narrower width of the posterior distribution compared to PolyChord, aligning with previous findings (Lemos et al. 2023; DK23; Li et al. 2023a). However, as expected, the results from the two sampling codes show consistency in terms of best-fit values.

In addition, we compared our cosmic shear results with those from the CMB analysis by the *Planck* satellite, using their baseline ΛCDM chains with the `Plik` likelihood from their most recent *Planck*-2018 results (Planck Collaboration VI 2020). More specifically, we used their constraints based on the auto power spectra of temperature (TT), of *E*-modes (EE), and their cross-power spectra (TE), excluding CMB lensing signals. An offset is evident between our results and those from *Planck*-2018. Adopting the Hellinger distance tension metric (Beran 1977; Heymans et al. 2021; DK23), we detect a 2.35σ tension in the constrained S_8 values. For the constrained parameter set (S_8, Ω_m), a similar level of tension, 2.30σ, was found using the Monte Carlo exact parameter shift method (Raveri et al. 2020; DK23).

Figure 5 presents our primary S_8 constraints and compares them with those from other contemporary cosmic shear surveys and the *Planck* CMB analysis. For ease of comparison, we show all three summary statistics for our fiducial results, while for other surveys, we display their headline values, as
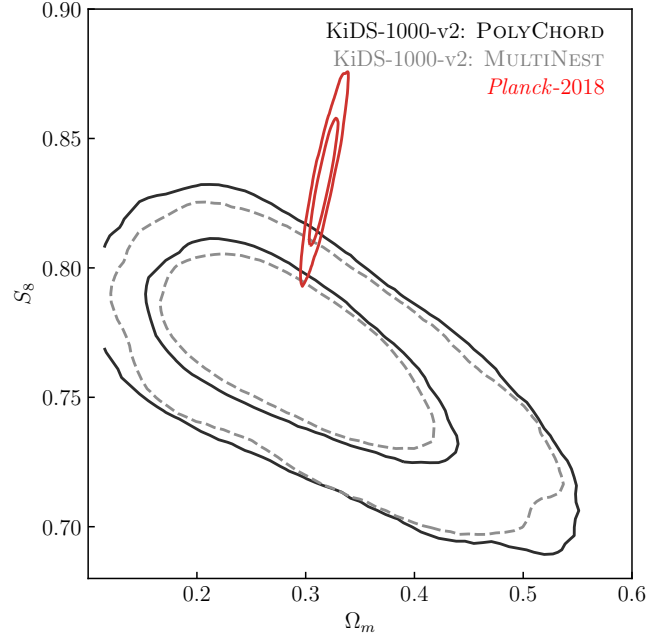
Fig. 4. Comparison of projected 2D posterior distributions for the parameters Ω_m and S_8 as derived from our fiducial setups using two sampling codes – PolyChord (solid black line) and MultiNest (dashed grey line) – against the *Planck*-2018 results (solid red line). The contours correspond to the 68% and 95% credible intervals and are smoothed using a Gaussian kernel density estimation (KDE) with a bandwidth scaled by a factor of 1.5, made possible by the `ChainConsumer` package (Hinton 2016).

per their preferred summary statistics. Overall, our results align well with those from all major contemporary cosmic shear surveys.

We note that our fiducial analysis pipeline is similar to the DK23 Hybrid pipeline with one notable difference: while DK23 included a free neutrino parameter, we kept the total neutrino mass fixed. DK23 showed that this additional degree of freedom in the cosmological parameter space can slightly increase the projected marginal S_8 values relative to an analysis with a fixed neutrino mass. However, since we refer to their MAP and PJ-HPD results in Fig. 5, the comparison should not be influenced by these projection effects (for more details, refer to the discussion in DK23).

It is interesting to note that our fiducial results align almost identically with the KiDS-1000-v1 re-analysis conducted by DK23, who used the A21 redshift calibration. This alignment arises from a balance of several effects in our analysis. Our improved shear calibration tends to increase S_8, while the enhanced vdB22 redshift calibration tends to lower it. Moreover, thanks to our enhanced empirical corrections for PSF leakages, our S_8 constraints are less affected by changes in the small scale cut used in measuring two-point correlation functions. The shifts we observe are roughly two times smaller than those in the KiDS-1000-v1 re-analysis conducted by DK23, which we discuss in more detail in Sect. 5.3.2. This helps reconcile the minor difference between our results and those of A21.

5.2. Impact of shear biases

The primary aims of this study are to assess the impact of higher-order shear biases on the final parameter constraints and to develop a methodology for effectively addressing shear
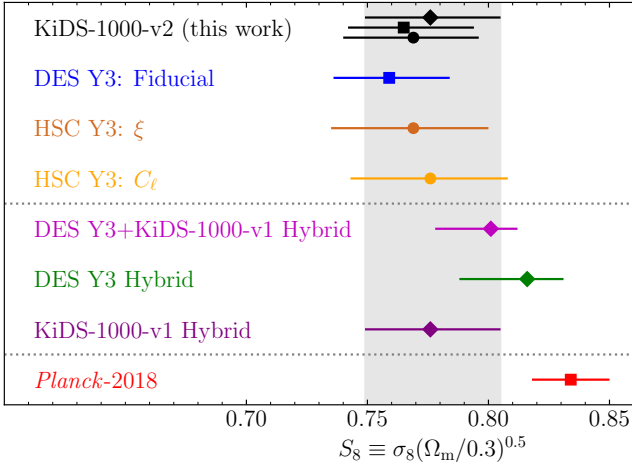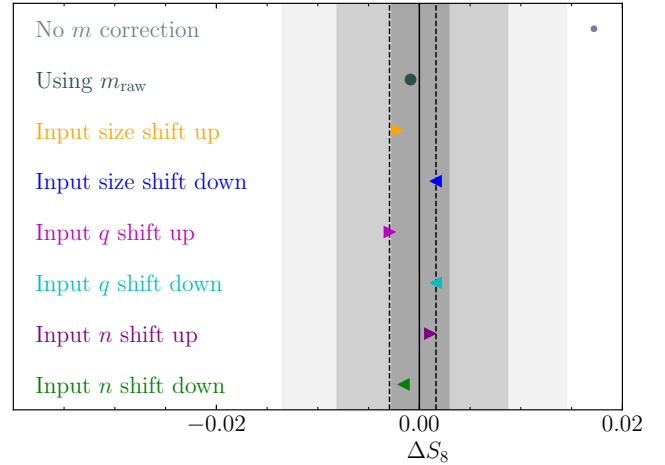
Fig. 5. Marginalised constraints on S_8 derived from our fiducial analysis with POLYCHORD, compared with those from other contemporary cosmic shear surveys and the *Planck* CMB analysis. Three sections, separated by dotted horizontal lines, indicate results of different origins. The first section includes results from individual cosmic shear surveys with their own analysis pipelines. The second section presents results from a collaborative effort between the DES and KiDS teams, who built a hybrid pipeline for analysing the data from both groups (DK23). The final section displays results from the *Planck* CMB analysis. Different labels are used for different statistical methods: the diamond represents results using the MAP and PJ-HPD statistics, the square denotes the mean-marginal statistics, and the circle shows the maximum-marginal statistics. The error bars correspond to the 68% credible intervals.

Fig. 6. Shifts in best-fit values of S_8 under different residual shear bias scenarios. The shift, ΔS_8, is calculated as $\Delta S_8 = S_8^{\rm test} - S_8^{\rm fiducial}$, where $S_8^{\rm test}$ represents the best-fit values in the test scenarios determined by a local minimisation method that uses the best-fit values from the fiducial analysis ($S_8^{\rm fiducial}$) as a starting point. The grey shaded regions represent different percentiles of the credible intervals derived from our fiducial POLYCHORD run. From the innermost to the outermost region, these percentiles are 6.8%, 20.4%, and 34%, corresponding to 0.1, 0.3, and 0.5 fractions of the reported sampling uncertainties. The dashed lines display the maximum shifts encountered in the six sets of morphology test simulations. These maximum shifts are used as the additional uncertainties in the reported best-fit values to account for the systematic uncertainties arising from shear calibration.

calibration uncertainties. Both of these aims can be achieved by examining the shifts in the constrained cosmological parameters resulting from different shear bias scenarios. As discussed in Sect. 4 and Appendix C, the residual shear biases have only a minor effect on the measured data vector. This allows us to determine the shifts in the best-fit values of the constrained parameters using a local minimisation algorithm, such as the Nelder–Mead method (Nelder & Mead 1965). These shifts in the best-fit values indicate the additional uncertainties stemming from systematic uncertainties in shear calibration.

Figure 6 shows shifts in our primary S_8 constraints for different residual shear bias scenarios. For comparison, we also include a shaded region denoting different levels of PJ-HPD credible intervals, as derived from our fiducial POLYCHORD chain. Apart from the extreme case where no shear calibration is applied, all other residual shear bias scenarios result in shifts less than 10% of the initial sampling uncertainties. Notably, neglecting the higher-order correction for the shear-interplay effect and uncertainties in PSF modelling results in a negligible shift of only -0.03σ (labelled "Using $m_{\rm raw}$" in the figure). This finding reinforces the reliability of previous KiDS cosmic shear analyses, which did not consider these higher-order effects.

The S_8 shifts, resulting from the input morphology test simulations, indicate additional systematic uncertainties in our shear bias calibration. To generate these test simulations, we changed the input values for three morphological parameters of the adopted Sérsic profile: the half-light radius (labelled "size" in the figure), axis ratio (labelled "q"), and the Sérsic index (labelled "n"). The adjustments were based on the fitting uncertainties reported by Griffith et al. (2012), from whose catalogue we derived the input morphology for our simulations (refer to Sect. 2.1.2 of L23). For simplicity, we shifted all galaxies in the same direction for each test simulation, implicitly assum-

ing that the fitting uncertainties stem from a coherent bias in that direction. This means that our test results represent the most extreme scenario. To consider both directions, we adjusted the input values in both positive and negative directions, leading to a total of six test simulations. Further details regarding the generation and comparison of these test simulations can be found in Appendix C.

We observe that shifts in the input galaxy axis ratio lead to the most significant changes in S_8: a -0.10σ shift for increased input axis ratio and a $+0.06\sigma$ shift for decreased input axis ratio. This behaviour aligns with our expectations for the *lens*fit code employed in our analysis. As it incorporates prior information on measured galaxy ellipticities during its Bayesian fitting process, it is more sensitive to changes in the distributions of sample ellipticities.

These S_8 shifts, obtained from the test simulations, provide a quantitative measure of the potential impact of inaccuracies in the input morphology and the sensitivity of the *lens*fit code to the underlying sample morphology distributions. When presenting the S_8 constraints, we accounted for these systematic uncertainties by including the maximum shifts in the reported uncertainties. In other words, we considered the shifts corresponding to the changes in input axis ratio (represented as dashed lines in Fig. 6), from the six sets of test simulations, as additional systematic uncertainties. These are reported alongside the original statistical uncertainties from the main sampling chain. It should be noted that these additional systematic uncertainties are specific to the SKiLLS image simulations and the *lens*fit shape measurement code used in our analysis. To reduce these uncertainties, future advancements in shear measurements should focus on improving the realism of image simulations and enhancing the robustness of the shear measurement algorithm.

5.3. Impact of altering inference setups

Although our main updates revolve around the shear measurement and calibration, we also implemented several modifications to the cosmological inference pipeline, drawing upon recent developments from DK23. As such, it is beneficial to conduct some extended runs with various setup configurations.

For these test runs, we employed MULTINEST as our sampling code, as it operates approximately five times faster than POLYCHORD, but at the cost of underestimating the width of the posterior distributions and thus the reported uncertainties by about 10%. However, the best-fit values from MULTINEST are not biased (as evident in Fig. 4). Thus, comparisons made using MULTINEST will yield conservative but unbiased results.

5.3.1. Priors for the NLA model

We began by testing the prior for the NLA model. As discussed in Sect. 4, our fiducial analysis implemented a redshift-independent NLA model with a narrow flat prior for the amplitude parameter A_{IA}. This model, motivated by the work of Fortuna et al. (2021a), serves as an alternative to the uninformative broad prior previously used. To investigate the impact of this change on our final results, we performed two additional runs: one employing a redshift-independent NLA model with a broad A_{IA} prior ranging from $[-6, 6]$, in line with KiDS-1000-v1 analyses, and another allowing for a redshift-dependent IA amplitude, namely, the NLA-z variant. The redshift evolution is modelled using a power law of the form $[(1 + z)/(1 + 0.62)]^{\eta_{IA}}$, with priors of $[-5, 5]$ for both A_{IA} and η_{IA}, in line with DK23.

Figure 7 presents a comparison of the posterior distributions obtained from the different NLA prior setups, and Table 4 lists the point estimates for the critical S_8 parameter. We see consistent constraints on S_8 across all setups. The constrained A_{IA} under our narrower prior setup also aligns with those from the broad priors, albeit spanning a narrower range due to the constrained prior range, validating the prior range used in our fiducial analysis. Additionally, we observe that the η_{IA} parameter is not constrained by the data, suggesting that the use of the NLA-z model may not be necessary for current weak lensing analyses.
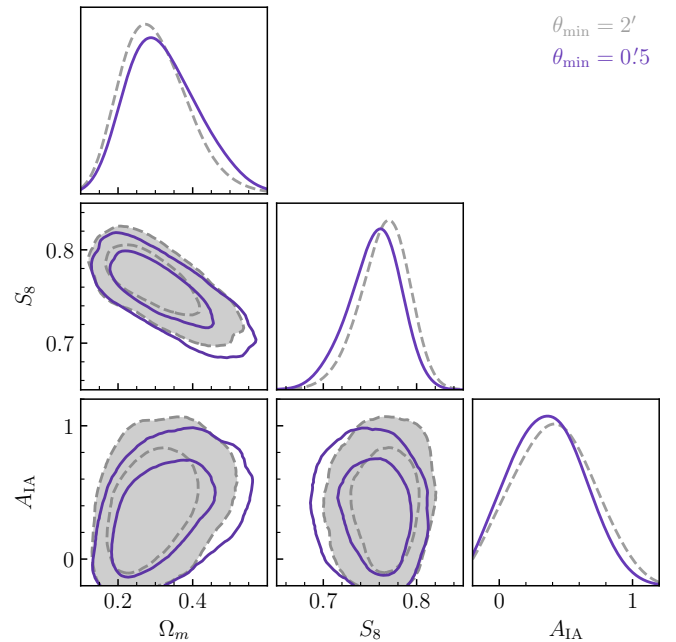
5.3.2. Different scale cuts

In our fiducial analysis, we adopted a scale cut for the measured data vectors, ranging from 2′ to 300′, as suggested by DK23. This is a change from the KiDS-1000-v1 analyses, which used a range of $0.5 < \theta < 300'$. A re-analysis of KiDS-1000-v1 with this new scale cut by DK23 led to a $0.7-0.8\sigma$ increase in the S_8 constraint. Using mock analyses, they found that this offset could arise from noise fluctuations 23% of the time.

In light of the updates to our shear measurement, we revisited this test. Interestingly, as shown in Fig. 8, we observe a smaller difference between the two scale cuts than what was reported by DK23. Specifically, we observe shifts of -0.17σ, -0.40σ, and -0.31σ, corresponding to the MAP and PJ-HPD, mean marginal, and maximum marginal summary statistics, respectively (refer to Table 4 for exact values).

We attribute this increased robustness against small scale fluctuations to our improved empirical corrections of the PSF leakages into shear measurement. This is supported by Figs. 2 and 3, where we see that the shear signals measured from the KiDS-1000-v2 catalogues exhibit overall smaller systematic errors. We note that Giblin et al. (2021) performed a mock test using the two-point correlation function and identified a change
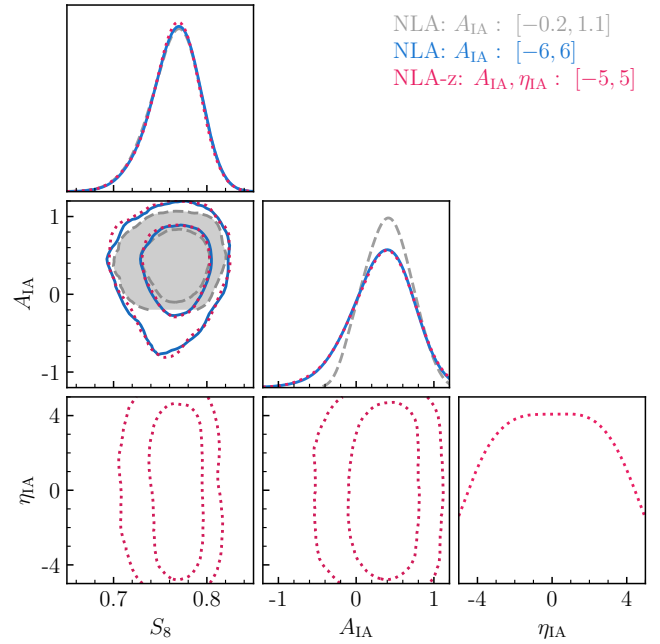
Fig. 7. Comparison of projected posterior distributions for the parameters S_8, A_{IA}, and η_{IA}, derived from three different NLA prior setups. The contours correspond to the 68% and 95% credible intervals and are smoothed using Gaussian KDE with a bandwidth scaled by a factor of 1.5.

Fig. 8. Comparison of projected posterior distributions for the parameters Ω_m, S_8, and A_{IA}, derived from two different scale cuts. The contours correspond to the 68% and 95% credible intervals and are smoothed using Gaussian KDE with a bandwidth scaled by a factor of 1.5.

of less than 0.1σ in the S_8 constraints when the detected PSF residuals were incorporated into the KiDS-1000-v1 mock data. Nevertheless, it is plausible that these systematic effects have a more significant influence on COSEBIs, given their use of more sophisticated weighting functions (Schneider et al. 2010). To quantify the improvements brought about by the updated

Table 4. Point estimates for S_8 from different inference setups.

Setups	MAP and PJ-HPD			Mean marginal		Maximum marginal	
	χ^2	S_8	ΔS_8	S_8	ΔS_8	S_8	ΔS_8
Fiducial: POLYCHORD	62.7	$0.776^{+0.029+0.002}_{-0.027-0.003}$	–	$0.765^{+0.029}_{-0.023}$	–	$0.769^{+0.027}_{-0.029}$	–
Fiducial: MULTINEST	62.7	$0.774^{+0.027}_{-0.023}$	0.00σ	$0.765^{+0.028}_{-0.017}$	0.00σ	$0.771^{+0.023}_{-0.027}$	0.00σ
NLA: A_{IA}: $[-6, 6]$	62.7	$0.774^{+0.034}_{-0.020}$	0.02σ	$0.766^{+0.028}_{-0.020}$	0.03σ	$0.770^{+0.023}_{-0.026}$	-0.01σ
NLA-z: A_{IA}, η_{IA}: $[-5, 5]$	62.7	$0.776^{+0.022}_{-0.025}$	0.10σ	$0.765^{+0.027}_{-0.020}$	0.00σ	$0.770^{+0.023}_{-0.025}$	-0.05σ
Scales: $0.5 < \theta < 300'$	64.5	$0.769^{+0.014}_{-0.036}$	-0.17σ	$0.755^{+0.029}_{-0.022}$	-0.40σ	$0.763^{+0.022}_{-0.030}$	-0.31σ
KiDS-1000-v1 setups	64.9	$0.771^{+0.017}_{-0.029}$	-0.13σ	$0.755^{+0.026}_{-0.020}$	-0.44σ	$0.759^{+0.022}_{-0.024}$	-0.50σ

Notes. "Fiducial: POLYCHORD" denotes our headline results, which are the same as those presented in Table 3. "Fiducial: MULTINEST" represents the same parameter setup as our fiducial analysis, but employs the MULTINEST sampling code. We used this as the reference to assess test results because all test runs utilise the MULTINEST code for increased speed. When comparing the Fiducial: MULTINEST results with the primary POLYCHORD results, we can conclude that the best-fit values obtained from MULTINEST are unbiased. The relative shift in S_8, denoted as ΔS_8, is calculated by comparing the best-fit values from the test runs to the reference Fiducial: MULTINEST run. The ΔS_8 values are expressed as a fraction of σ, which signifies the standard deviation of estimates from the test run. We calculated ΔS_8 for different summary statistics separately for consistency. For MAP and PJ-HPD results, we also present the best-fit χ^2 values. For comparison, the best-fit χ^2 values from A21 and vdB22 are 82.2 and 63.2, respectively.

shear measurements regarding the robustness of the COSEBIs, a similar mock analysis based on the COSEBIs statistic is warranted. We consider this an important topic for future study. For the current analysis, the test results simply affirm the robustness of our primary S_8 constraints.

5.3.3. KiDS-1000-v1 setups

To draw a direct comparison with the KiDS-1000-v1 results and evaluate the impact of our improved shear measurements and calibration, we performed a test run using the same inference pipeline and parameter priors as in the KiDS-1000-v1 analyses conducted by A21 and vdB22. The differences compared to our fiducial analysis setup include: measurements from scales of 0.5 to $300'$, use of the older version of HMCODE, sampling with the MULTINEST code, and a broad A_{IA} prior ranging from $[-6, 6]$. As shown in Fig. 9, our test results are well aligned with the outcomes of the analyses by A21 and vdB22. Notably, our new results show an increase in the S_8 value relative to vdB22, bringing it closer to the result obtained by A21.

We re-emphasise that our redshift calibration aligns with that of vdB22, who expanded the redshift calibration sample to more than double the size used by A21 (see Appendix B for details). This means that our redshift-related selection function closely mirrors that used in the vdB22 sample. However, due to changes in the weighting and selection scheme between the KiDS-1000-v2 catalogue and the KiDS-1000-v1 catalogue, our sample cannot be considered as directly comparable to theirs.

To provide a more quantitative understanding of the sample differences among the three analyses, we compared the effective number density of the source sample in our analysis to those used in A21 and vdB22. The differences for each tomographic bin are 9.6%, 9.8%, 6.1%, 10.6%, and 2.8% when compared to A21; and -1.8%, -1.3%, -0.7%, 0.7%, and 3% when compared to vdB22. Here, positive values signify an increase, while negative values denote a decrease. The differences between our catalogue and that of A21 stem from both shear measurement and redshift calibration, whereas the difference between ours and that of vdB22 arises mainly from the shear measurement, as we used the same SOM for the "gold" selection (see Appendix B). As such, comparing our results directly with those of vdB22 can provide
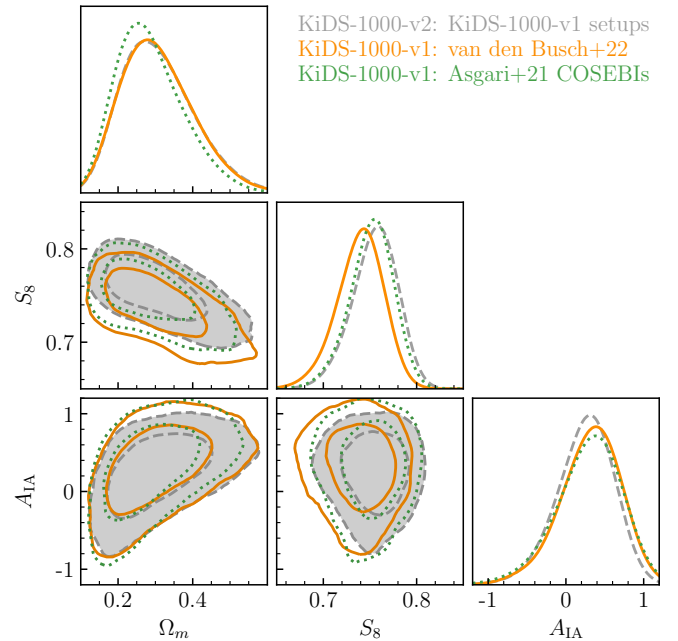
Fig. 9. Comparison of projected posterior distributions for parameters Ω_m, S_8, and A_{IA} from our analysis (dashed grey lines) based on the KiDS-1000-v2 catalogue, to those from vdB22 (solid orange lines) and A21 (dotted green lines), both of which are based on the KiDS-1000-v1 catalogues. The cosmological inference pipeline and parameter priors are identical across all three analyses presented here. In terms of measurements, vdB22 and A21 used the same shear measurements and calibration, while vdB22 and our analysis share the same redshift calibration. The contours correspond to the 68% and 95% credible intervals and are smoothed using Gaussian KDE with a bandwidth scaled by a factor of 1.5.

clearer insights into the impact of our improvements in shear measurements. It is also worth noting that the increased effective number density in high redshift bins compared to vdB22 is largely due to the increased weighting of faint objects in the updated version of *lens*fit code. However, this comes at the cost of increased sample ellipticity dispersion, with a maximum

increase of 6% found in the fifth bin. These subtle differences in the source catalogues change the noise properties of the samples. Consequently, even with perfect calibration in each study, we would not expect to derive identical cosmological constraints from each analysis.

Interestingly, the increase in number density from A21 to vdB22, and in this work, does not significantly reduce the marginalised uncertainties of the final cosmological parameters. This can be largely attributed to the fact that the majority of the constraining power in the KiDS analysis comes from the high redshift bins, as illustrated in Fig. 7 of A21, whereas our increase in number density is most pronounced in the lower redshift bins. Additionally, changes in the redshift distributions, due to alterations in the redshift calibration sample, could further impact the final constrained uncertainties, as demonstrated in Table 3 of vdB22. Lastly, due to the intricate degeneracy among nuisance and cosmological parameters, caution should be used when inferring that an increased number density will directly lead to a reduction in the marginalised uncertainties of specific parameters.

6. Summary

We have conducted a cosmic shear analysis using the KiDS-1000-v2 catalogue, which is an updated version of the public KiDS-1000(-v1) catalogue with respect to shear measurements and calibration. Under the assumption of a spatially flat ΛCDM cosmological model, we derived constraints on $S_8 = 0.776^{+0.029+0.002}_{-0.027-0.003}$ based on the MAP and PJ-HPD summary statistics. The second set of uncertainties was incorporated to account for the systematic uncertainties within our shear calibration. The mean-marginal and maximum-marginal values obtained from the same sampling chain are $0.765^{+0.029}_{-0.023}$ and $0.769^{+0.027}_{-0.029}$, respectively. Our results are consistent with earlier results from KiDS-1000-v1 and other contemporary weak lensing surveys but show a $\sim 2.3\sigma$ level of tension with the *Planck* CMB constraints.

The main improvements in our analysis, relative to the KiDS-1000-v1 cosmic shear analyses, are attributed to the enhanced cosmic shear measurement and calibration. These enhancements were achieved through the updated version of the *lens*fit shape measurement code, a new empirical correction scheme for PSF contamination, and the newly developed SKiLLS multi-band image simulations, as detailed in L23. We verified the reliability of the new measurement via a series of catalogue-level null tests proposed by Giblin et al. (2021). The results indicate that the KiDS-1000-v2 catalogue shows overall better control over measurement systematics compared to the KiDS-1000-v1 catalogues. This improvement in reducing measurement systematics helps in reducing noise in small scale measurements, thereby enhancing the robustness of our cosmological parameter constraints against varying scale cut choices.

Our methodology for shear calibration largely aligns with the one detailed in L23, where we accounted for higher-order blending effects that arise when galaxies from different redshifts are blended, as well as the uncertainties in PSF modelling. However, when comparing the outcomes from the shear calibration with and without these higher-order adjustments, we find that these effects have a negligible impact on the present weak lensing analysis, a conclusion that is in line with the findings of Amon et al. (2022).

We recommend treating the statistical and systematic uncertainties from the shear calibration separately, given their distinct origins. The statistical uncertainties, which are determined by the simulation volume, can be reduced and incorporated into the

covariance matrix used for cosmological inference. On the other hand, systematic uncertainties, associated with the realism of image simulations and the sensitivity of the shape measurement algorithm, can be more effectively addressed when considered as residual shear biases post-calibration. Assuming these residual shear biases are small, a forward modelling approach, combined with a local minimisation method, can be used to estimate their impact on the final parameter constraints. In our analysis, these additional systematic uncertainties contribute roughly 8% of the final uncertainty on S_8. However, ongoing efforts to enhance shear measurement and calibration, such as increasing the realism of image simulations through Monte Carlo control loops (Refregier & Amara 2014) and leveraging new techniques such as METACALIBRATION/METADETECTION (Huff & Mandelbaum 2017; Sheldon & Huff 2017; Sheldon et al. 2020; Hoekstra et al. 2021) to improve measurement robustness against underlying sample properties, may well lead to a reduction in these additional systematic uncertainties.

In our fiducial analysis, we opted for a redshift-independent NLA model with a narrow flat prior for the IA amplitude parameter, A_{IA}, motivated by the work of Fortuna et al. (2021a). However, we also investigated two alternative scenarios: one with a broad A_{IA} prior for the redshift-independent NLA model, echoing the KiDS-1000-v1 analysis by A21, and the other the NLA-z variant, allowing for redshift evolution of the IA amplitude, as per the recent joint DES Y3+KiDS-1000 cosmic shear analysis (DK23). In all three scenarios, we find fully consistent constraints for S_8 and A_{IA}, which indicates that the impact of the variations is negligible in these scenarios. To better understand the IA signals and their impact on cosmic shear analyses, future tests need to implement more substantial variations in IA models, for instance the halo model formalism introduced by Fortuna et al. (2021a). Such an exploration would not only enhance our understanding of the measured IA signals, but also help mitigate correlations between nuisance parameters, thereby improving the precision of future cosmic shear analyses.

The second group covers those who have either made a significant contribution to the data products, or to the scientific analysis.

References

Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, PASJ, 70, S4
Amon, A., & Efstathiou, G. 2022, MNRAS, 516, 5355
Amon, A., Gruen, D., Troxel, M. A., et al. 2022, Phys. Rev. D, 105, 023514
Asgari, M., Heymans, C., Hildebrandt, H., et al. 2019, A&A, 624, A134
Asgari, M., Tröster, T., Heymans, C., et al. 2020, A&A, 634, A127
Asgari, M., Lin, C.-A., Joachimi, B., et al. 2021, A&A, 645, A104
Bacon, D. J., Refregier, A. R., & Ellis, R. S. 2000, MNRAS, 318, 625
Bacon, D. J., Massey, R. J., Refregier, A. R., & Ellis, R. S. 2003, MNRAS, 344, 673
Benítez, N. 2000, ApJ, 536, 571
Beran, R. 1977, Ann. Stat., 5, 445
Blandford, R. D., Saust, A. B., Brainerd, T. G., & Villumsen, J. V. 1991, MNRAS, 251, 600
Bridle, S., & King, L. 2007, New J. Phys., 9, 444
Chang, C., Jarvis, M., Jain, B., et al. 2013, MNRAS, 434, 2121
Dalal, R., Li, X., Nicola, A., et al. 2023, ArXiv e-prints [arXiv:2304.00701]
Dark Energy Survey Collaboration (Abbott, T., et al.) 2016, MNRAS, 460, 1270
Dawson, W. A., Schneider, M. D., Tyson, J. A., & Jee, M. J. 2016, ApJ, 816, 11
de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, Exp. Astron., 35, 25
DES and KiDS Collaboration (Abbott, T. M. C., et al.) 2023, Open J. Astrophys., 6, 36
Edge, A., Sutherland, W., Kuijken, K., et al. 2013, The Messenger, 154, 32
Efstathiou, G., & Lemos, P. 2018, MNRAS, 476, 151
Elahi, P. J., Welker, C., Power, C., et al. 2018, MNRAS, 475, 5338
Euclid Collaboration (Martinet, N., et al.) 2019, A&A, 627, A59
Fenech Conti, I., Herbonnet, R., Hoekstra, H., et al. 2017, MNRAS, 467, 1627
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Fischbacher, S., Kacprzak, T., Blazek, J., & Refregier, A. 2023, JCAP, 2023, 033
Fortuna, M. C., Hoekstra, H., Joachimi, B., et al. 2021a, MNRAS, 501, 2983
Fortuna, M. C., Hoekstra, H., Johnston, H., et al. 2021b, A&A, 654, A76
Giblin, B., Heymans, C., Asgari, M., et al. 2021, A&A, 645, A105
Girardi, L., Groenewegen, M. A. T., Hatziminaoglou, E., & da Costa, L. 2005, A&A, 436, 895
Griffith, R. L., Cooper, M. C., Newman, J. A., et al. 2012, ApJS, 200, 9
Handley, W. J., Hobson, M. P., & Lasenby, A. N. 2015a, MNRAS, 450, L61
Handley, W. J., Hobson, M. P., & Lasenby, A. N. 2015b, MNRAS, 453, 4384
Hartlap, J., Hilbert, S., Schneider, P., & Hildebrandt, H. 2011, A&A, 528, A51
Heymans, C., Van Waerbeke, L., Bacon, D., et al. 2006, MNRAS, 368, 1323
Heymans, C., Tröster, T., Asgari, M., et al. 2021, A&A, 646, A140
Hikage, C., Oguri, M., Hamana, T., et al. 2019, PASJ, 71, 43
Hilbert, S., Hartlap, J., White, S. D. M., & Schneider, P. 2009, A&A, 499, 31
Hildebrandt, H., Choi, A., Heymans, C., et al. 2016, MNRAS, 463, 635
Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, MNRAS, 465, 1454
Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2020, A&A, 633, A69
Hildebrandt, H., van den Busch, J. L., Wright, A. H., et al. 2021, A&A, 647, A124
Hinton, S. R. 2016, J. Open Sour. Softw., 1, 00045
Hirata, C. M., & Seljak, U. 2004, Phys. Rev. D, 70, 063526
Hoekstra, H., Herbonnet, R., Muzzin, A., et al. 2015, MNRAS, 449, 685
Hoekstra, H., Viola, M., & Herbonnet, R. 2017, MNRAS, 468, 3295
Hoekstra, H., Kannawadi, A., & Kitching, T. D. 2021, A&A, 646, A124
Howlett, C., Lewis, A., Hall, A., & Challinor, A. 2012, JCAP, 2012, 027
Hoyle, B., Gruen, D., Bernstein, G. M., et al. 2018, MNRAS, 478, 592
Hu, W. 1999, ApJ, 522, L21
Huff, E., & Mandelbaum, R. 2017, ArXiv e-prints [arXiv:1702.02600]
Huterer, D. 2002, Phys. Rev. D, 65, 063001
Jarvis, M. 2015, Astrophysics Source Code Library [record ascl:1508.007]
Jarvis, M., Bernstein, G., & Jain, B. 2004, MNRAS, 352, 338
Joachimi, B., Mandelbaum, R., Abdalla, F. B., & Bridle, S. L. 2011, A&A, 527, A26
Joachimi, B., Cacciato, M., Kitching, T. D., et al. 2015, Space Sci. Rev., 193, 1
Joachimi, B., Lin, C. A., Asgari, M., et al. 2021, A&A, 646, A129
Johnston, H., Georgiou, C., Joachimi, B., et al. 2019, A&A, 624, A30
Kaiser, N. 1992, ApJ, 388, 272
Kaiser, N., Wilson, G., Luppino, G. A., et al. 2000, ApJL, submitted [arXiv:astro-ph/0003338]

Kannawadi, A., Hoekstra, H., Miller, L., et al. 2019, A&A, 624, A92
Kitching, T. D., Miller, L., Heymans, C. E., van Waerbeke, L., & Heavens, A. F. 2008, MNRAS, 390, 149
Kitching, T. D., Paykari, P., Hoekstra, H., & Cropper, M. 2019, Open J. Astrophys., 2, 5
Köhlinger, F., Joachimi, B., Asgari, M., et al. 2019, MNRAS, 484, 3126
Kohonen, T. 1982, Biol. Cybern., 43, 59
Krause, E., Fang, X., Pandey, S., et al. 2021, ArXiv e-prints [arXiv:2105.13548]
Kuijken, K., Heymans, C., Hildebrandt, H., et al. 2015, MNRAS, 454, 3500
Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, A&A, 625, A2
Lagos, C. D. P., Tobar, R. J., Robotham, A. S. G., et al. 2018, MNRAS, 481, 3573
Lemos, P., Weaverdyck, N., Rollins, R. P., et al. 2023, MNRAS, 521, 1184
Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473
Li, S.-S., Kuijken, K., Hoekstra, H., et al. 2021, A&A, 646, A175
Li, X., Zhang, T., Sugiyama, S., et al. 2023a, ArXiv e-prints [arXiv:2304.00702]
Li, S.-S., Kuijken, K., Hoekstra, H., et al. 2023b, A&A, 670, A100
Lima, M., Cunha, C. E., Oyaizu, H., et al. 2008, MNRAS, 390, 118
MacCrann, N., Becker, M. R., McCullough, J., et al. 2022, MNRAS, 509, 3371
Mandelbaum, R. 2018, ARA&A, 56, 393
Mandelbaum, R., Lanusse, F., Leauthaud, A., et al. 2018, MNRAS, 481, 3170
Massey, R. 2010, MNRAS, 409, L109
Massey, R., Hoekstra, H., Kitching, T., et al. 2013, MNRAS, 429, 661
Masters, D., Capak, P., Stern, D., et al. 2015, ApJ, 813, 53
McCarthy, I. G., Schaye, J., Bird, S., & Le Brun, A. M. C. 2017, MNRAS, 465, 2936
Mead, A. J., Peacock, J. A., Heymans, C., Joudaki, S., & Heavens, A. F. 2015, MNRAS, 454, 1958
Mead, A. J., Heymans, C., Lombriser, L., et al. 2016, MNRAS, 459, 1468
Mead, A. J., Brieden, S., Tröster, T., & Heymans, C. 2021, MNRAS, 502, 1401
Melchior, P., & Viola, M. 2012, MNRAS, 424, 2757
Miller, L., Kitching, T. D., Heymans, C., Heavens, A. F., & van Waerbeke, L. 2007, MNRAS, 382, 315
Miller, L., Heymans, C., Kitching, T. D., et al. 2013, MNRAS, 429, 2858
Miralda-Escude, J. 1991, ApJ, 380, 1
Nelder, J. A., & Mead, R. 1965, Comput. J., 7, 308
Paulin-Henriksson, S., Amara, A., Voigt, L., Refregier, A., & Bridle, S. L. 2008, A&A, 484, 67
Perivolaropoulos, L., & Skara, F. 2022, New Astron. Rev., 95, 101659
Planck Collaboration VI. 2020, A&A, 641, A6
Preston, C., Amon, A., & Efstathiou, G. 2023, MNRAS, 525, 5554
Raveri, M., Zacharegkas, G., & Hu, W. 2020, Phys. Rev. D, 101, 103527
Refregier, A., & Amara, A. 2014, Phys. Dark Univ., 3, 1
Refregier, A., Kacprzak, T., Amara, A., Bridle, S., & Rowe, B. 2012, MNRAS, 425, 1951
Rhodes, J. D., Massey, R. J., Albert, J., et al. 2007, ApJS, 172, 203
Samuroff, S., Bridle, S. L., Zuntz, J., et al. 2018, MNRAS, 475, 4524
Samuroff, S., Mandelbaum, R., Blazek, J., et al. 2023, MNRAS, 524, 2195
Schneider, P., van Waerbeke, L., & Mellier, Y. 2002, A&A, 389, 729
Schneider, P., Eifler, T., & Krause, E. 2010, A&A, 520, A116
Secco, L. F., Samuroff, S., Krause, E., et al. 2022, Phys. Rev. D, 105, 023515
Semboloni, E., Hoekstra, H., Schaye, J., van Daalen, M. P., & McCarthy, I. G. 2011, MNRAS, 417, 2020
Sheldon, E. S., & Huff, E. M. 2017, ApJ, 841, 24
Sheldon, E. S., Becker, M. R., MacCrann, N., & Jarvis, M. 2020, ApJ, 902, 138
Singh, S., Mandelbaum, R., & More, S. 2015, MNRAS, 450, 2195
Tanaka, M., Coupon, J., Hsieh, B.-C., et al. 2018, PASJ, 70, S9
Toyozumi, H., & Ashley, M. C. B. 2005, PASA, 22, 257
Tröster, T., Asgari, M., Blake, C., et al. 2021, A&A, 649, A88
Troxel, M. A., & Ishak, M. 2015, Phys. Rep., 558, 1
van Daalen, M. P., Schaye, J., Booth, C. M., & Dalla Vecchia, C. 2011, MNRAS, 415, 3649
van Daalen, M. P., McCarthy, I. G., & Schaye, J. 2020, MNRAS, 491, 2424
van den Busch, J. L., Hildebrandt, H., Wright, A. H., et al. 2020, A&A, 642, A200
van den Busch, J. L., Wright, A. H., Hildebrandt, H., et al. 2022, A&A, 664, A170
Van Waerbeke, L., Mellier, Y., Erben, T., et al. 2000, A&A, 358, 30
Wittman, D. M., Tyson, J. A., Kirkman, D., Dell'Antonio, I., & Bernstein, G. 2000, Nature, 405, 143
Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. 2020, A&A, 640, L14
Zuntz, J., Paterno, M., Jennings, E., et al. 2015, Astron. Comput., 12, 45

Appendix A: Shear bias in two-point statistics

When calibrating the shear measurements in the two-point correlation function, it is usually assumed that the correlations involving the shear bias can be ignored, which includes correlations between different tomographic and spatial angular bins. This simplification leads to the following relationship between the true correlation function of cosmic shear in tomographic bins i, j, denoted as ξ^{ij}, and the measured signal $\hat{\xi}^{ij}$:

$$
\begin{aligned}
\hat{\xi}^{ij}(\theta) &= \langle \hat{\gamma}^i(\theta')\,\hat{\gamma}^j(\theta'+\theta)\rangle \\
&= \langle [1+m^i(\theta')]\,[1+m^j(\theta'+\theta)]\,\gamma^i(\theta')\,\gamma^j(\theta'+\theta)\rangle \\
&= \langle [1+m^i(\theta')]\,[1+m^j(\theta'+\theta)]\rangle\,\xi^{ij}(\theta) \\
&\simeq (1+\overline{m^i})(1+\overline{m^j})\,\xi^{ij}(\theta),
\end{aligned}
\tag{A.1}
$$

where $\overline{m^i}$ is estimated by averaging over all sources in a given tomographic bin i, and we use $\langle\cdot\rangle$ to denote the correlation function. We also assumed that the shear bias is independent of the underlying shear to simplify the equation. The result of Eq. (A.1) allows us to average the multiplicative biases over all the galax-

ies in a given tomographic bin to mitigate the individual noisy bias estimation.

However, in principle, the shear bias can be scale dependent due to spatial fluctuations in source density (e.g. Samuroff et al. 2018). With SKiLLS, we can directly examine these correlations by measuring the shear bias in the two-point estimators. We measured the shear correlation function in the SKiLLS mock catalogue using Eq. (2). Since we know the true $\xi^{ij}_+(\theta) = \gamma^2_{\mathrm{input}}$ in simulations, where γ_{input} is the amplitude of the constant input shear, we can estimate the shear bias in the two-point correlation function directly by comparing the measured $\hat{\xi}^{ij}_+$ to the input ξ^{ij}_+ following Eq. (A.1).

Figure A.1 shows the difference between the shear biases with and without considering its correlations, defined as $\Delta m_\xi \equiv \langle [1+m^i(\theta')]\,[1+m^j(\theta'+\theta)]\rangle - (1+\overline{m^i})(1+\overline{m^j})$. It shows that the difference is negligible across all scales and tomographic bins, in agreement with the statistical uncertainties of our shear calibration, which are represented by the shaded regions. These findings confirm that we can neglect the correlations between shear biases in the current KiDS weak lensing analysis.
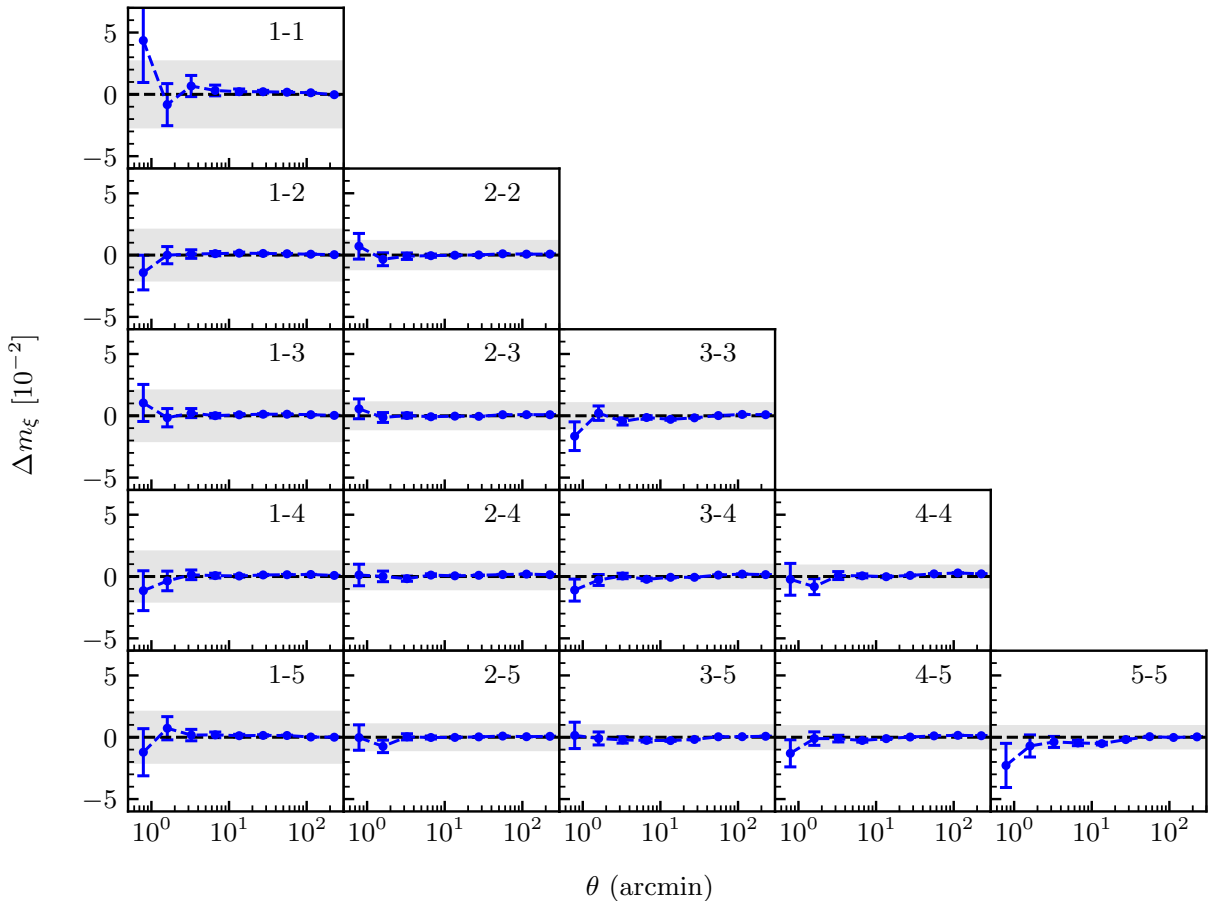
Fig. A.1. Two-point correlations between the multiplicative shear biases. The correlation is estimated as $\Delta m_\xi \equiv \hat{\xi}^{ij}_+/\gamma^2_{\mathrm{input}} - (1+\overline{m^i})(1+\overline{m^j})$. The 15 panels represent the different combinations of the five redshift bins utilised in our cosmic shear analysis. The shaded regions within each panel denote the statistical uncertainties of our shear calibration for each tomographic bin, as outlined in Table 1.

Appendix B: SOM redshift calibration

This appendix provides information on the redshift calibration reference sample and SOM configurations used in our analysis. For a more comprehensive overview and validation of the SOM redshift calibration method in the KiDS analysis, we refer to Wright et al. (2020), Hildebrandt et al. (2021), and vdB22.

We employed the fiducial spectroscopic sample described in vdB22 as our calibration reference sample. This sample comprises spectroscopic redshift estimates (spec-zs) from various spectroscopic surveys that overlap with KiDS fields, enabling us to assign KiDS photometric measurements to objects in the reference sample. In cases where an object had multiple spectroscopic measurements, vdB22 defined a specific hierarchy to select the most reliable redshift estimates based on the quality of the measurements. For further details on the adopted spectroscopic samples and the compilation procedure, readers are referred to Appendix A of vdB22.

For our calibration, we used a 101×101 hexagonal SOM trained on the r-band magnitude and 36 colours derived from the PSF-matched, list-driven nine-band $ugriZYJHK_s$ photometry from the KiDS+VIKING surveys. This SOM is identical to the fiducial SOM constructed in vdB22. We segregated the reference and target samples into the trained SOM cells separately for each tomographic bin, allowing us to create comparable groupings between the spectroscopic and photometric sources in each bin. During this process, we further categorised the original SOM cells using a hierarchical cluster analysis implemented by the "hclust" function within the R Stats Package[12] to increase the number of galaxies per grouping. We adopted the same number of clusters per bin (4000, 2200, 2800, 4200, and 2000) as (Wright et al. 2020), who determined these numbers using simulations produced by van den Busch et al. (2020).

To mitigate the effects of photometric noise and the incompleteness of the reference sample, we applied an additional selection step to the SOM groupings. We excluded any grouping where the mean spectroscopic redshift of the reference sample $\overline{z_{\mathrm{spec}}}$ and the mean photometric redshift of the target sample $\overline{z_{\mathrm{B}}}$ exhibited a significant discrepancy, defined as $|\overline{z_{\mathrm{spec}}} - \overline{z_{\mathrm{B}}}| > 5\sigma_{\mathrm{mad}}$. Here, σ_{mad} represents the normalised median absolute deviation of all SOM groupings, which we calculated to be 0.122 in our case. This step allowed us to define the KiDS "gold" sample, which we used to compute the redshift distributions and perform the cosmic shear analysis.

Appendix C: Systematic uncertainties from the shear calibration

In this appendix, we outline our approach to addressing the systematic uncertainties arising from shear calibration. Our methodology involves two primary steps: In Sect. C.1, we quantify the potential residual biases after implementing our simulation-based shear calibration. In Sect. C.2, we propagate these systematic uncertainties into the final uncertainties of the estimated cosmological parameters.

We propose a separate accounting of the shear calibration uncertainties, as it is considered more accurate and informative than the traditional approach, which uses nominal shear calibration uncertainties that are deliberately overestimated to encompass potential systematic uncertainties arising from shear calibration. Our approach clearly illustrates the extent to which

the final cosmological parameters of interest are influenced by these systematic uncertainties from shear calibration.

Furthermore, as mentioned in Sect. 3.2, these systematic uncertainties have fundamentally different origins from the statistical uncertainties incorporated in the covariance matrix. They represent the fundamental limitations of current simulation-based shear calibration methods. The limitations inherent in these systematic uncertainties cannot be eliminated by merely increasing the scale of image simulations. However they can be mitigated by empirically enhancing the realism of the image simulations, for example, using the Monte-Carlo Control Loop method (Refregier & Amara 2014), or by improving the robustness of the shear measurement algorithm, such as the METACALIBRATION/METADETECTION method (Huff & Mandelbaum 2017; Sheldon & Huff 2017; Sheldon et al. 2020; Hoekstra et al. 2021).

### C.1. Quantifying residual shear biases with sensitivity analysis

Residual biases may persist after simulation-based shear calibration due to imperfect alignment between simulations and data, as elucidated by K19. These discrepancies pose challenges for shear calibration methods dependent on image simulations and underscore the need for re-weighting simulations to more closely align with the data. However, given that intrinsic galaxy properties in real data are unknown, this re-weighting process relies on noisy measured properties, rendering it vulnerable to calibration selection biases as discussed by FC17. The uncertainties linked with the measured properties cause galaxies to be intermixed among defined bins, leading to the up-weighting or down-weighting of certain galaxies. As a result, even if the re-weighted sample aligns with the data in terms of the distribution of measured properties, it does not ensure identicality in terms of intrinsic properties. In other words, shear biases can still vary between two samples with identical distributions of apparent measured properties. Our aim is to quantify these residual biases and incorporate them into the final uncertainties of cosmological parameters.

The SKiLLS multi-band image simulations used in this analysis incorporate several enhancements, informed by insights gathered from previous KiDS simulation studies (FC17; K19). These improvements include: reproducing variations in star density, PSF, and noise background across the KiDS footprint; incorporating faint galaxies down to an r-band magnitude of 27 to account for correlated noise from undetected objects (e.g. Hoekstra et al. 2017); including realistic clustering from N-body simulations to address blending effects (e.g. K19); and adopting an end-to-end approach for photo-z estimation to account for photo-z measurement uncertainties. These improvements augment the robustness of the shear biases estimated from SKiLLS against various observational conditions.

In an investigation on the propagation of observational biases in shear surveys, Kitching et al. (2019) demonstrated that the measured shear power spectrum is, to first order, predominantly influenced by the mean of the multiplicative bias field across a survey. This suggests that if the shear bias estimated from simulations accurately reflects the mean value of the targeted sample, the shear calibration will be robust enough for KiDS-like cosmic shear analyses. Therefore, we conclude that potential residual biases related to observational conditions have negligible influence on our shear calibration, and we focused on systematic uncertainties arising from galaxy morphology uncertainties, specifically the assumed Sérsic profile and its parameters derived

---

[12] https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/hclust
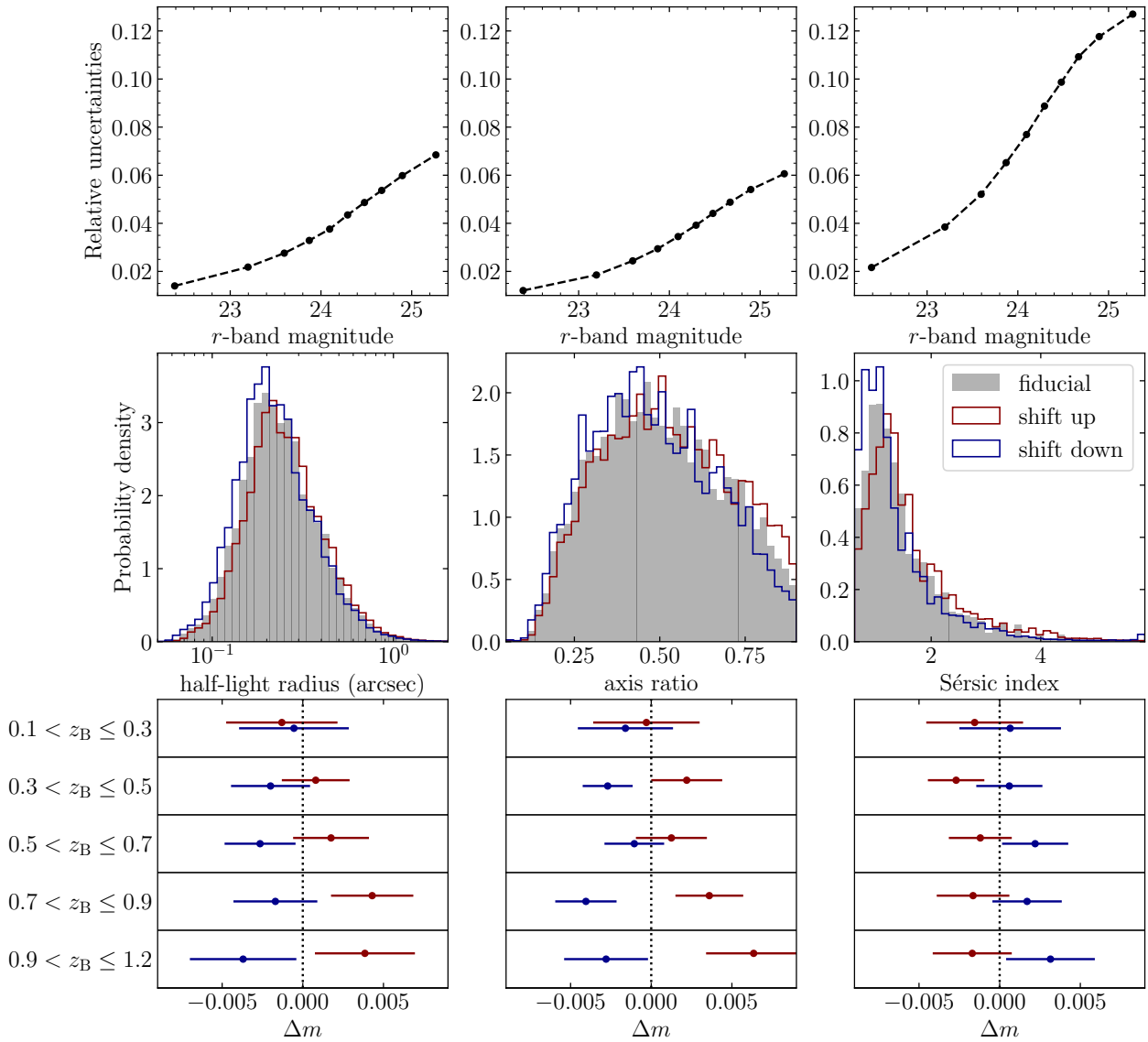
Fig. C.1. Changes in input morphological parameter values for assessing residual biases after simulation-based shear calibration. From left to right, the order is half-light radius, axis ratio, and Sérsic index. *Top panels*: Relationship between relative fitting uncertainties and the r-band magnitude, as measured from the catalogues of Griffith et al. (2012). The values shown are calculated using the median values of the binned samples. *Middle panels*: Overall distributions of input morphological parameters. Comparisons are drawn among the fiducial (grey shades), the test sample with input values increased by an amount corresponding to the relative uncertainties (red lines), and the test sample with input values decreased by the same amount of relative uncertainties (blue lines). *Bottom panels*: Shifts in multiplicative biases in tomographic bins resulting from changes in input morphological parameter values. Both fiducial and test simulations are re-weighted using the same dataset via the method detailed in Sect. 3.1.

from *Hubble* Space Telescope observations (Griffith et al. 2012). For a model-fitting shape measurement code like *lens*fit, these galaxy morphology uncertainties are the main sources of residual shear biases after implementing the simulation-based shear calibration.

The deviation from the Sérsic profile is challenging to address for the current SKiLLS simulations, as our copula-based learning algorithm requires a parameterised model for its application. However, the Sérsic model has been validated as sufficient for KiDS-like analyses by K19, who used the same morphology catalogue as our work. Thus, we focus on the measurement uncertainties of the Sérsic parameters: half-light radius, axis ratio, and Sérsic index. We first examined the fitting uncertainties reported by Griffith et al. (2012) to assess the

accuracy of these parameters in our input catalogue. We find that the median relative uncertainties for these parameters are a smooth function of galaxy magnitude, as shown in the top panels of Fig. C.1. This allows us to capture these correlations through simple linear interpolation.

We interpreted these relative uncertainties as indicators of the systematic uncertainties in our input morphology. We assumed the most extreme scenarios, in which these measured statistical uncertainties are all caused by a coherent bias in the same direction. Consequently, we adjusted all galaxies in our input sample in the same direction, with the amplitude of the adjustment determined based on their r-band magnitude using a simple linear interpolation of the measured median correlations. We examined shifts towards both larger and smaller values and

considered the three Sérsic parameters separately. This resulted in six test simulations corresponding to the six different sets of variations in input morphology parameter values. The input parameter distributions for these test simulations, as shown in the middle panels of Fig. C.1, are compared to the distributions of the fiducial simulations. A clear shift of the entire distribution is evident, suggesting that our test simulations represent the most extreme scenarios in which the measured statistical uncertainties are coherently biased in the same direction, a situation that is unlikely in reality. Therefore, the residual biases we identified from these test simulations provide a conservative estimate.

We applied the same data analysis procedures to the test simulations as we did to the fiducial simulations, including shear and redshift estimates. We also followed the same re-weighting procedure for the test simulations as for the fiducial simulations, ensuring that the calibration selection biases are also captured. The differences in shear biases between these test simulations and our fiducial simulation are illustrated in the bottom panels of Fig. C.1. The small differences indicate that the residual shear biases, after implementing our fiducial shear bias calibration, are insignificant.

## C.2. Propagating residual shear biases with forward modelling

Accurately incorporating the systematic uncertainties from shear calibration into the covariance matrix presents a challenge, as residual shear biases directly scale the data vector, as shown in Eq. (A.1). A more direct approach is to assess the shift in the measured shear signal caused by the residual shear biases and evaluate how these data vector shifts influence the constrained

cosmological parameters. Given the minor residual shear biases illustrated in Fig. C.1 and the unchanged covariance, it is not necessary to reiterate the sampling of the posterior distributions for each shift. Instead, we can implement a local minimisation algorithm to find nearby best-fit values for each shift, using starting points from the fiducial sampling chain. The range of these new best-fit values, each associated with a shift, indicates the additional systematic uncertainties introduced by the residual shear biases.

This approach naturally integrates with our existing cosmological inference method, as outlined in Sect. 4, which already requires an additional local optimisation step to refine the best-fit values identified by the sampling code. We simply replicated this optimisation step, using the original best-fit value as the starting point and the shifted likelihood to determine the best-fit values associated with various alterations in measured signals. The variability in these test best-fit values provides an expanded credible region for the inferred parameters, thereby representing the systematic uncertainties from shear calibration. We included these additional uncertainties when presenting the point estimates of our primary parameters (see Sect. 5.2 for details).

Appendix D: Contour plots for all free parameters

In this appendix, we provide two supplementary contour plots that display the posterior distributions of all twelve free parameters from our fiducial analyses, as produced by both the POLYCHORD and MULTINEST sampling codes. The overall concordance between the results generated by POLYCHORD and MULTINEST is evident.
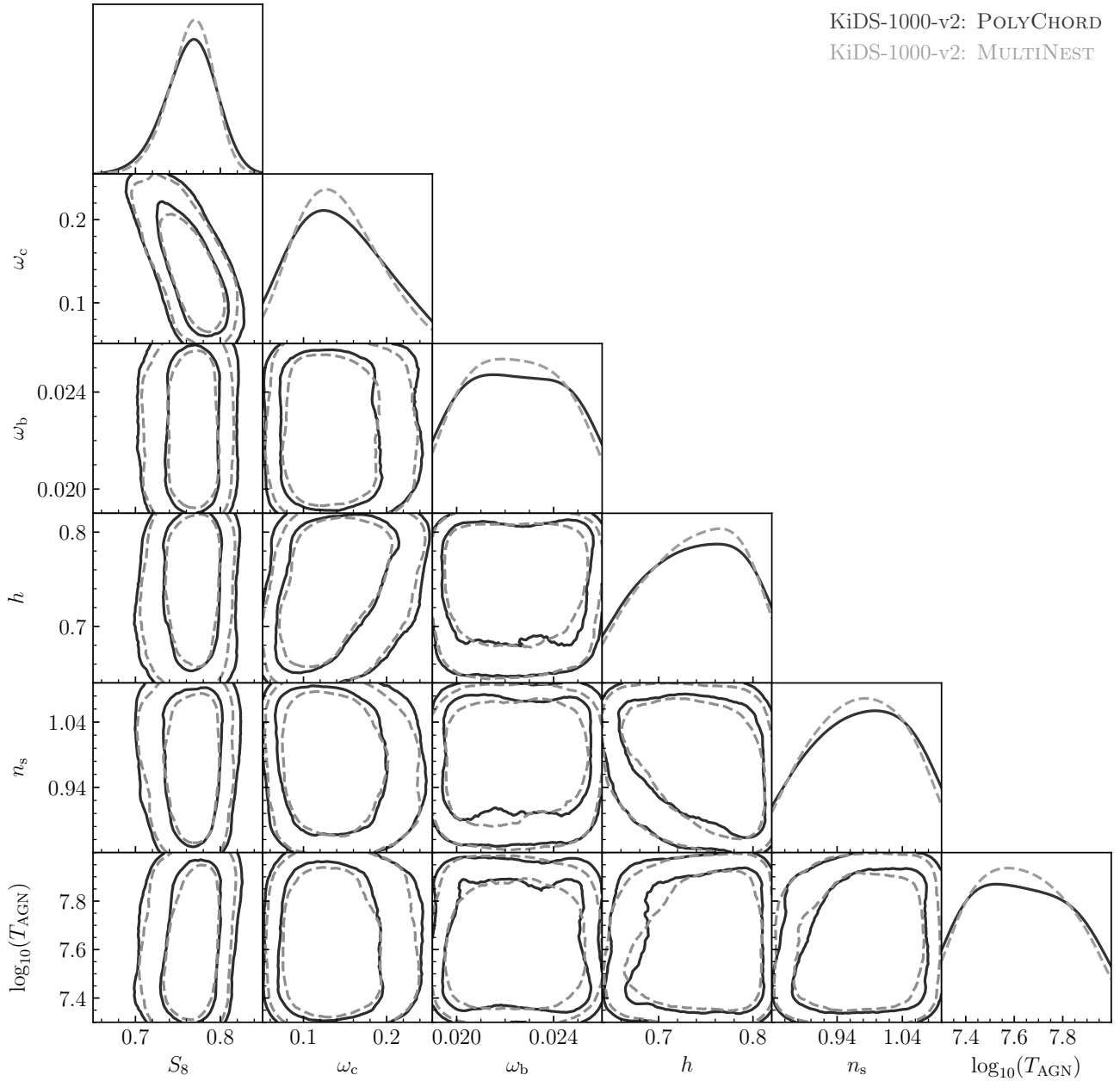
Fig. D.1. Posterior distributions of cosmological and baryonic parameters from our fiducial analyses, as generated by POLYCHORD (solid black lines) and MULTINEST (dashed grey lines). The contours represent the 68% and 95% credible intervals and are smoothed using a Gaussian KDE with a bandwidth scaled by a factor of 1.5. We note that S_8 is the only parameter that our data robustly constrain.
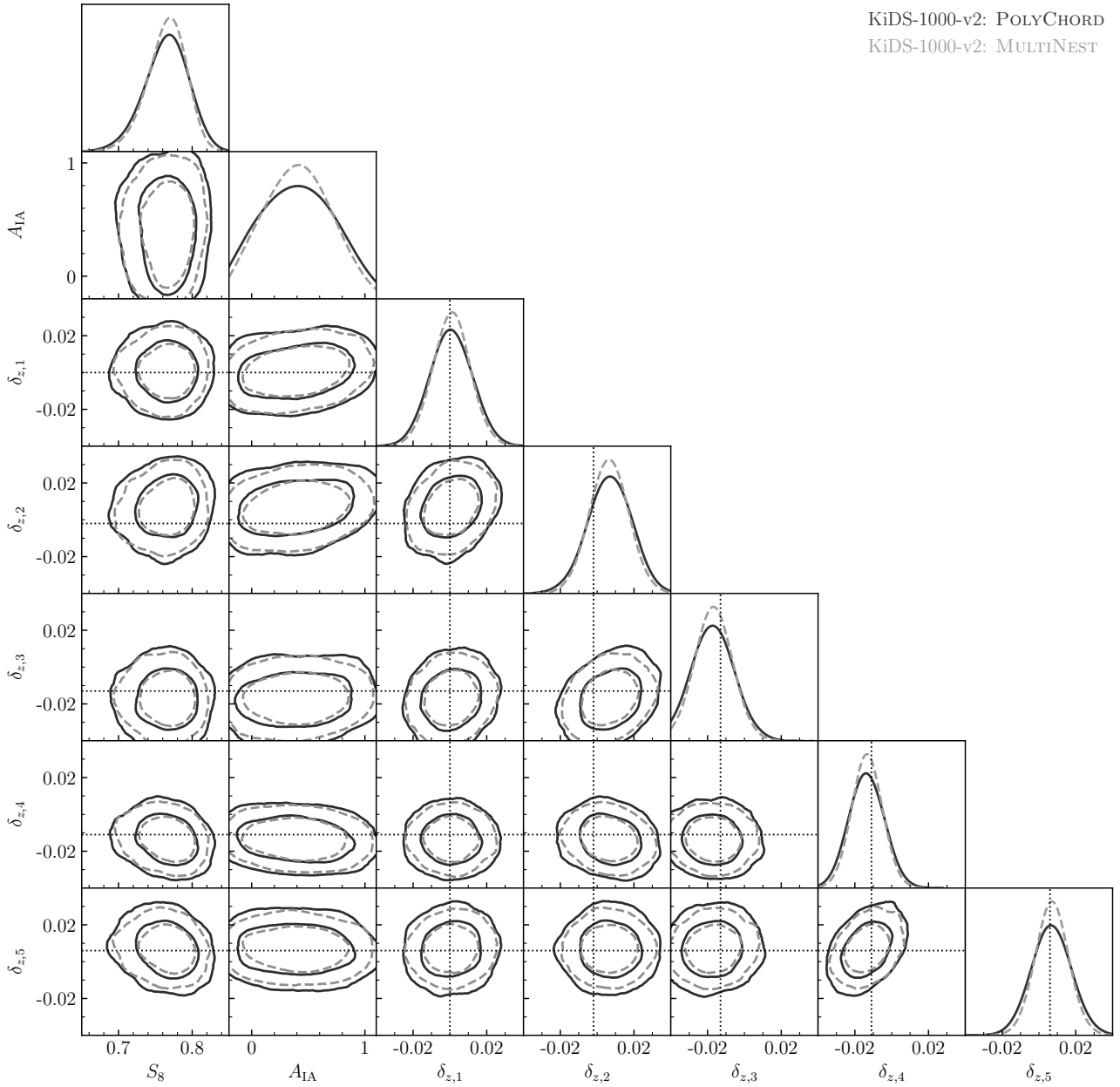
Fig. D.2. Posterior distributions of S_8 and nuisance parameters from our fiducial analyses, as generated by PolyChord (solid black lines) and MultiNest (dashed grey lines). The contours represent the 68% and 95% credible intervals and are smoothed using a Gaussian KDE with a bandwidth scaled by a factor of 1.5. We note that the Gaussian priors we have set, as outlined in Table 2, strongly influence the redshift offset parameters. The dotted lines represent the central values of these Gaussian priors.